IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Group Art Unit: 1804
Examiner: Suzanne E. Ziska, Ph.D.

In re PATENT APPLICATION of

Applicant : Christopher R. BEBBINGTON)
Appln. No. : 08/633,013)
Filed : April 16, 1996)
For : RECOMBINANT DNA EXPRESSION VECTORS)
Atty. Dkt. : CARPR 0037D2)

RESPONSE

November 25, 1996

FEE Enclosed: $ NONE
Please charge any further fee to Dep. Acct. 19-3700

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

In response to the Official Action dated July 24, 1996, Applicant respectfully submits that this Response is being timely filed under the Next-Business-Day Rule (November 24, 1996 fell on a Sunday). Kindly consider the Response as follows:

1. At the top of page 2 of the Official Action of July 24, 1996, claims 6, 7 and 11-18 stand rejected as being indefinite in the recitation of "recombinant".

Applicant respectfully submits that the term "recombinant" has a well defined, art recognized meaning. Furthermore, claims to the vector identical to the instant transformed host cell claims have already been allowed. For the Examiner's convenience the following table presents the allowed claims from parent application 08/300,063:

CLAIM NUMBER	CLAIM
11	A recombinant expression vector comprising the promoter, enhancer and complete 5' untranslated region including the first intron of the hCMV-MIE gene operably linked to a heterologous coding sequence.
12	The recombinant expression vector according to claim 11 further comprising a restriction site to facilitate insertion of the heterologous coding sequence.
13	The recombinant expression vector according to claim 11 wherein the promoter, enhancer and complete 5' untranslated region including the first intron of the hCMV-MIE gene are linked directly to the heterologous coding sequence.
14	The recombinant expression vector according to claim 11 wherein the vector further includes the hCMV MIE gene's includes a translational initiation signal.
15	The recombinant expression vector according to claim 14 wherein the translational initiation signal includes the sequence 5'-GTCACCGTCCTTGACACCATG-3'.
16	The recombinant expression vector according to claim 14 wherein the translational initiation signal includes the sequence 5'-CCATGG-3'.

2. On page 2 of the Official Action of July 24, 1996, claims 6, 7 and 11-18 stand rejected as being anticipated by Whittle et al. (Protein Engineering, 1987).

Applicant respectfully submits that the Examiner appears to have mistakenly read the publication year of Whittle et al. as 1985 (see page 2, line 20 of the Official Action of July 24, 1996). Review of the Whittle et al. publication indicates that

the actual publication year was 1987. Applicant respectfully submits that the Whittle et al. publication does not constitute prior art under any section of 35 USC 102. Although the copy of the reference provided by the Examiner does not set forth a publication date, page 505 of the publication in the last sentence after the "Reference" section, indicates that the paper was received for review on **October 8, 1987 and revised in December 8, 1987**. As the priority date for the instant application is July 23, 1987, the Whittle et al. reference does not constitute prior art under 35 USC § 102 because the article was not even submitted for review at the priority date of the instant application.

Applicant respectfully directs the Examiner to British Application No. GB 8717430 to which priority is claimed under 35 USC § 119. This document was submitted on December 1, 1995 in parent application 08/300,063 in which the issue fee was paid on May 22, 1996. For the Examiner's convenience, a photocopy of said application is attached as Exhibit A. Applicant respectfully requests official acknowledgement of the priority document submitted in parent application 08/300,063 and withdrawal of the above rejection.

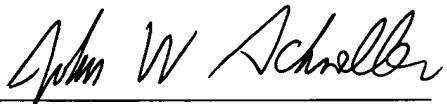
3. In view of the foregoing Remarks, Applicant respectfully requests the Examiner to reconsider and withdraw her rejections of the claims. Should the Examiner feel that an interview would

expedite the prosecution of this application, she is invited to call the undersigned at her convenience.

The Commissioner is hereby authorized to charge any additional fees which may be required for this amendment, or credit any overpayment to Deposit Account No. 19-3700.

An extension of time is being filed with this Response. In the event that the required fee for a one (1) month extension of time is not attached, or fees which may be required in addition to that requested in a petition for an extension of time, the Commissioner is requested to grant a petition for that extension of time which is required to make this response timely and is hereby authorized to charge any fee for such an extension of time or credit any overpayment for an extension of time to Deposit Account No. 19-3700 and to notify the undersigned accordingly.

Respectfully submitted,



John W. Schneller
(Registration No. 26,031)
SPENCER & FRANK
Suite 300 East
1100 New York Ave., N.W.
Washington, D.C. 20005-3955
Telephone: (202) 414-4000
Telefax : (202) 414-4040

JWS:MST:dvb

PCT/GB 88/00060

30 September 19
09 19

THE PATENT OFFICE
STATE HOUSE
66—71 HIGH HOLBORN
LONDON WCIR 4TP

Exhibit A

REC'D 2 6 OCT 1988
WIPO PCT

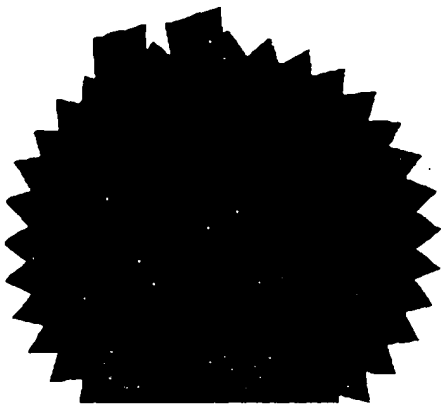
I, the undersigned, being an officer duly authorised in accordance with Section 62(3) of the Patents and Designs Act 1907, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or the inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words, "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Witness my hand this
19th day of SEPTEMBER 1988



23 JULY 1987

PATENTS ACT 1977

PATENTS FORM No. 1/77 (Revised 1982)

(Rules 16, 19)

1987
17430

The Comptroller
The Patent Office
25 Southampton Buildings
London, WC2A 1AY

REQUEST FOR GRANT OF A PATENT

8717430

THE GRANT OF A PATENT IS REQUESTED BY THE UNDERSIGNED ON THE BASIS OF THE PRESENT APPLICATION

I Applicant's or Agent's Reference *(Please insert if available)* G03217/PA140

II Title of Invention Recombinant DNA Product and Processes Using It.

III Applicant or Applicants *(See note 2)*

Name (First or only applicant) Celltech Limited

Country United Kingdom State ADP Code No.

Address 216 Bath Road, Slough, Berkshire,
United Kingdom.

Name (of second applicant, if more than one)

Country State

Address

IV Inventor (see note 3)

(a) [redacted]
or
(b) A statement on Patents Form No 7/77 [redacted] will be furnished

V Name of Agent (if any) *(See note 4)* **CARPMAELS & RANSFORD** ADP CODE NO

VI Address for Service (See note 5) **43, BLOOMSBURY SQUARE LONDON, WC1A 2RA**

VII Declaration of Priority (See note 6)

Country	Filing date	File number
–	–	–

VIII The Application claims an earlier date under Section 8(3), 12(6), 15(4), or 37(4) (See note 7)

Recombinant DNA Product and Processes Using It

The present invention relates to DNA sequences which cause spontaneous high copy number incorporation of vector DNA into a host cell and to the uses thereof in recombinant DNA technology.

The main aim of workers in the field of recombinant DNA technology is to achieve as high a level of production as possible of a particular polypeptide. This is particularly true of commercial organisations who wish to exploit the use of recombinant DNA technology to produce polypeptides which naturally are not very abundant.

Generally, the application of DNA technology involves the cloning of a gene encoding the desired polypeptide, placing the cloned gene in a suitable expression vector, transfecting a host cell line with the vector, and culturing the transfected cell line to produce the polypeptide. Optionally, the process may include vector amplification stages in an attempt to raise the production level. These steps are now well known and for the most part can be operated satisfactorily. However, there is still much uncertainty as to how much polypeptide will in the end be produced. It is almost impossible to predict whether any particular vector or cell line or combination thereof will lead to a useful level of production.

There are in general two factors which significantly affect the amount of polypeptide produced by a transfected cell line. The first factor relates to the efficiency at which the cloned gene is transcribed and translated in the host cell. The present application is not primarily concerned with this factor.

The second factor relates to the number of copies of the gene present in the transfected cell. If there is a large number of copies, an increased level of production can be expected. There have been

a number of proposals for increasing copy number. The most commonly used is generally known as vector amplification, and is best exemplified by the DHFR system.

DHFR is dihydrofolate reductase, an enzyme which confers on a cell line the ability to grow in the absence of nucleosides in the medium. In a typical DHFR-based amplification system, a dhfr$^-$ cell line is transfected with a gene encoding DHFR and a gene encoding the desired polypeptide. The transfected cell line is then grown in medium lacking nucleosides. Cells which survive may contain both the DHFR gene and the desired gene. Surviving cells are then cultured in media containing increasing concentrations of methotrexate (MTX), a compound which binds to DHFR, thereby inhibiting its action. The surviving cells have amplified levels of the DHFR gene and concomitantly amplified levels of the gene encoding the desired polypeptide.

While amplification systems have been relatively successful in increasing copy number, they are far from perfect, in that they require a number of rounds of culturing, which is very time consuming. There is therefore a need for a system whereby the copy number of a desired gene in a transfected cell line can be increased without the need for laborious amplification procedures,

A further problem with presently-known vector amplification systems is that initial transfectants containing low copy numbers of the vector may not produce sufficient product for detection. Thus, identification of clones for subsequent amplification may be difficult. There is therefore a need for a system which enables transfected cell lines to be identified more easily.

The present invention is based on the discovery

that the use of a particular vector to transfect a cell line led to the production of a transfected cell line having a surprisingly and unexpectedly high vector copy number, without the need to carry out any amplification procedures. It is nonetheless possible to carry out amplification in addition, to further increase vector copy number.

The vector which led to this discovery is the vector pSVLGS.1. The structure of this plasmid is shown in Figure 1. The plasmid is based on the vector pCT54 (Emtage et al., PNAS-USA, 80, 3671-3675, 1983), and comprises the EcoRI-BamHI fragment thereof. The remainder of the vector, going from the EcoRI site to the BamHI site, comprises the SV40 late region promoter and a minigene encoding glutamine synthetase (GS). The GS minigene comprises the complete coding sequence, a single intron and approximately 2kb of 3'-flanking DNA spanning two presumed sites of polyadenylation. The preparation of this vector is described in detail in International Patent Application No. PCT/GB87/00039.

The sequences in the pCT54 vector derived from plasmid pBR322 and the SV40 late region promoter have both been used in many systems, without giving rise to any unexpected increase in copy number. Moreover, other vectors have been produced which contain the GS coding sequence, but not the intron, of the GS minigene. Such vectors have been used for transfection without achieving spontaneously high vector copy number. It has now also been shown that spontaneously high vector copy numbers can be achieved using a vector lacking the PvuI-BamHI fragment of pCT54. It is therefore believed that the DNA sequence(s) in the vector which gives rise to the surprisingly and unexpectedly high copy number is located:

(i) in the intron;

(ii) in the 3' flanking DNA;

(iii) in a region bridging the SV40 portion and the GS coding portion;

(iv) in a region bridging one or the other end of the intron and the coding sequence;

(v) in a region bridging the coding sequence and the 3' flanking DNA;

(vi) in the 5'-untranslated region derived from the GS genomic DNA;

(vii) in the region of the 5'-untranslated region which is a cloning artefact; or

(viii) in any combination of the above.

This DNA sequence(s) is herein termed a "spontaneous high copy number sequence".

Almost the entire sequence of the vector pSVLGS.1 is known and is shown in Figure 2. The areas which correspond to those regions set out as (i) to (vii) in the paragraph above are marked. Work is at present being carried out to identify which particular region(s) is responsible for the ability of the vector to transfect a cell line with high copy number and to elucidate the mechanism by which the high copy number is spontaneously obtained. These experiments will merely be a matter of routine for the man skilled in the art and will lead to the identification of the exact sequence of the region(s) in question.

According to a first aspect of the present invention, there is provided: the DNA sequence(s) from the vector pSVLGS.1 which causes spontaneous high copy number incorporation of vector DNA into a host cell; or any DNA or RNA sequence which hybridises thereto under high stringency conditions; or any analog thereof. The spontaneous high copy number sequences according to this aspect of the present invention are hereinafter referred to as SVGS

The inventors conjecture that the SHCSs may comprise a pair of small inverted repeats which enable the formation of an intracellular molecular intermediate in gene amplification comprising two complete vector sequences in inverted relationship.

It has also been conjectured by the present inventors that the SHCSs correspond to sequences found around the break points in tandem arrays of amplified genes or to the sequences of hypervariable mini-satellite sequences. However, it is certainly not the Applicants desire to be limited in any way to these conjectured rationalisations.

If the present inventors' conjectures are correct then the SHCSs of the first aspect of the present invention will include sequences derived from the region around break points in amplified arrays of genes and sequences derived from hypervariable mini-satellite sequences. The SHCSs may also include sequences found in repeated sequences of mammalian genome, such as the "Alu" repeats, which may form sites at which recombination events can readily take place. An alternative or additional hypothesis is that the SHCSs comprise or include a mammalian origin of replication.

According to a second aspect of the present invention, there are provided vectors, in particular expression vectors, containing an SHCS. Preferably, the vector contains two such sequences but in inverted relationship. Alternatively, there is provided a pair of similar vectors containing a single SHCS, in each of which the SHCS is in the opposite orientation to its orientation in the other vector. It is believed that the use of such a vector or pair of vectors will enable a large inverted duplication to arise by homologous recombination within the host cell and hence induce amplification.

Preferably, the vector includes not only the SHCS but also a gene encoding a selectable marker. Such a vector will be of use in transforming a cell line in which the gene encoding the selectable marker is either absent or not expressed. A particularly suitable marker is GS. The use of GS as a selectable marker permits the survival of only those transfected cells which express a certain minimum level of GS which permits resistance to a certain level of methionine sulphoximine (MSX, a GS inhibitor). This is in contrast to other selection procedures (e.g. typically used selection procedures for DHFR or guanine phosphoribosyl transferase (GPT) genes) in which there is a less stringent requirement for efficient expression of the selected gene. It has been found that, using GS as a selectable marker, the frequency at which transfectants are identified after transfection with GS encoding genes is substantially lower than the frequency obtained by using a DHFR, GPT or neomycin-resistance gene as the selectable marker, since only a sub-set of transfectants (i.e. those which express higher than average levels of GS) can survive. Thus the use of GS as a selectable marker enables the selection of high copy number transfectants without the need to carry out any amplification stages. Nonetheless, the SHCSs could also be used in combination with a DHFR encoding sequence and using MTX as the selection agent.

Preferably, in expression vectors according to the second aspect of the present invention, the coding sequence is placed under the control of a very strong promoter to direct expression of the coding sequence. Advantageously, the promoter-containing fragment also includes sequences which allow efficient translation of the mRNA from the coding

sequence. In a particularly preferred embodiment, the coding sequence is placed under the control of a sequence comprising the promoter-enhancer and complete 5' untranslated region of the major immediate early gene of human cytomegalovirus (hCMV-MIE).

Preferably, the CMV-derived sequence includes both the first splice donor and splice acceptor site of the MIE gene and a sequence similar or identical to a concensus translation "start" signal.

The present invention also includes host cells transfected with the vectors according to the invention and processes for the production of a desired polypeptide by culturing such transfected cells.

According to a third aspect of the present invention, there is provided the use of the hCMV-MIE 5' untranslated region linked directly to the coding sequence for a desired polypeptide, for directing the translation of mRNA. It is believed that this 5' untranslated region is surprisingly efficient in directing mRNA translation.

The present invention is now described, by way of example only, with reference to the accompanying drawings, in which,

Figure 1 shows the structure of the vector pSVLGS.1;

Figure 2 shows the nucleotide sequence of the GS sequences in the vector pSVLGS.1, from the 5' EcoRI site to the HindIII site near the 3' end;

Figure 3 shows the structure of the vector pHT.1;

Figure 4 shows the complete sequence of the hCMV-MIE promoter enhancer region including the first intron and a modified translation "start" site; and

Figure 5 shows the structure of the vector pEE6HLC.HHC.GS.

Example 1

The construction of the vector pSVLGS.1 is shown in International Patent Application No. PCT/GB87/00039, a copy of which is enclosed with the present application. The International Application also shows the production of two other vectors comprising a GS coding sequence. The other two vectors are called pSV2.GS and pZIPGS. The structure of the pSVLGS.1 vector is shown in Figure 1 of this application and also in Figure 3A of the International Application. It can be seen from present Figure 1 that pSVLGS.1 includes a pCT54 derived section, the SV40 late promoter region, a GS coding sequence, including a single intron, and about 2 kb of 3' flanking DNA.

The structures of pSV2.GS and pZIPGS are shown in Figures 3B and 3C respectively of the International Application. In these vectors, the GS is encoded by a cDNA portion, lacking both the intron and the 3' flanking DNA. Moreover, the GS coding sequence is under the control of the SV40 early region promoter.

The results given in the International Application clearly show that the pSVLGS.1 vector is incorporated in very high copy number into a host cell merely on transfection, and in the absence of amplification. The differences in the structures of the pSVLGS.1 and the other two vectors, in particular the pSV2.GS vector, lead to the conclusion that the SHCS in the pSVLGS.1 vector is present in one or a combination of the seven regions enumerated above.

Example 2

A series of experiments was carried out to compare a number of vectors containing selectable markers. The three selectable markers used were GS, DHFR and GPT. In order to compare GS, DHFR and GPT selection, a cDNA encoding tissue inhibitor of metalloproteinase (TIMP) was used as a "reporter" gene. TIMP expression levels were studied from cell clones selected by each of the three methods.

The basic vector used in these experiments is pHT.1 which is shown in Figure 3. In this vector, the transcription unit used to direct TIMP expression contains the hCMV-MIE promoter including its complete 5' untranslated region fused by means of a single base change directly onto the NcoI site at the ATG representing the first amino acid of the TIMP coding sequence. This promoter-enhancer-leader fragment was made by adding an oligomer which recreates the entire 5' untranslated sequence to the Pst-1m fragment of hCMV. The complete sequence of the promoter-enhancer region and 5' translated region including the first intron and the modified translation "start" site is shown in Figure 4. At the 3' end of the TIMP cDNA fragment is the SV40 early polyadenylation signal. At the 3' end of this transcription unit is a unique BamHI site that was used to insert either i) the PvuI-BamHI fragment of pSVLGS.1 (which contains the GS minigene), ii) the PvuII-BamHI fragment from pSV2.dhfr (which contains the dhfr cDNA) or iii) the mouse metallothionein mMT-1 gene. Hence were derived three vectors pHT.1GS, pHT.1DHFR and pHT.1MT. In each case, both genes on the vector were in the same orientation. pHT.1GS and pHT.1DHFR were transfected into chinese hamster ovary (CHO) K1 and dhfr$^-$ CHO cells respectively and transfectants were selected

for i) resistance to 20 uM MSX (pHT.1GS) or ii) dhfr$^+$ phenotype (pHT.1DHFR). pHT.1MT was co-transfected with pEE6GPT (a vector containing the bacterial xanthine-guanine phosphoribosyl transferase (gpt) gene) into CHO-K1 cells and transfectants were selected for resistance to 5ug/ml mycophenolic acid in the presence of xanthine, hypoxanthine and thymidine. 24 colonies were isolated from each of the three transfections. These were grown-up and assayed for TIMP production. The results obtained are shown in Table 1.

TABLE 1

	pHT.1GS	pHT.1DHFR	pHT.1mMT
No of cell lines secreting TIMP	17/24	17/24	9/24

Several clones secreting the highest levels of TIMP from each transfection were grown to equivalent cell densities and TIMP secretion was assayed. The results are shown in Table 2.

TABLE 2

Clones	Timp levels ug ml^{-1}
GS Timp 10	5.5
GS Timp 14	5.5
GS Timp 15	9.0
GS Timp 19	8.5
dhfr Timp 1	0.8
dhfr Timp 3	0.7
dhfr Timp 4	0.58
dhfr Timp 5	1.5
mMT Timp 1	4.8
mMT Timp 2	3.0
mMT Timp 3	4.0

The GS TIMP clones produced 2-3 times more TIMP than mMT TIMP clones, and about 10 times more than dhfr TIMP clones.

The cell line secreting the most TIMP from each transfection was cloned by limiting dilution and the specific production rates for the best clone in each case were as follows:-

GS TIMP	19.12	10ug/10^6 cells/24 hours
dhfr TIMP	3.6	0.75ug/10^6 cells/24 hours
mMTTIMP	5.8	4ug.10^6 cells/24 hours

Selection using a vector including the SHCS dervived from the GS minigene and GS as a selectable marker allowed the identification of clones producing substantially higher levels of TIMP than were obtained using either of two alternative selectable markers DHFR or GPT.

Clones from the cell line GS TIMP 19 could also be selected for gene amplification by culturing in 500uM MSX and a cell line was obtained which secreted 100ug/10^6 cells/24 hours.

Example 3

An expression vector designed specifically for the expression of immunoglobulin genes in CHO cells was constructed. The structure of the expression vector, which is called pEE6HLC.HHC.GS, is shown in Figure 5. It contains the following DNA sequences: immunoglobulin light and heavy chain genes under the control of the human cytomegalovirus immediate early gene promoter and SV40 early gene polyadenylation signal; the GS minigene from pSVLGS.1 under the control of the SV40 late gene promoter; a bacterial origin of replication; and the amplicillin resistance gene. Following the introduction of plasmid DNA into CHO cells by calcium phosphate co-precipitation, colonies were isolated which were resistant 20uM MSX. These cell lines were subjected to a further

selection in 200uM MSX. Rates of antibody secretion were measured for each cell line and gene copy number was estimated from Southern blots of genomic DNA. The results are shown in Table 3.

The initial transfected cell lines have a copy number of at least fifty per cell and after a single round of selection for GS gene amplification this copy number increased to approximately five hundred. This increase in copy number is accompanied by a 00-20 fold increase in the rate of antibody secretion.

TABLE 3

Cell Line	Characteristics	Specific Production Rates ($g/10^6$cells/ 24 hours)	No. of vector copies/cell
36	Unamplified pool	0.077	50
36.5	Unamplified clone	0.125	50
36.1.i	Amplified pool	2.9	500
36.1.ii	Amplified pool	0.075	100
36.1 I	Amplified clone	3.19	500
36.1 J	Amplified clone	1.28	50

36.1 ii is the amplified pool after culturing for 2 months in the absence of MSX.

Example 4

In order to test whether the SHCS in the GS minigene can be used to obtain transfected clones expressing a linked gene more efficiently than vectors lacking the GS minigene, the following vectors were introduced into a dhfr$^-$ CHO cell line: 1) pSV2.dhfr, which is a widely used selectable marker in this cell line, conferring the ability to grow without added nucleosides; 2) pSV2dhfrGS3,

which contains the GS minigene (specifically the sequence between the Pvul and BamHI sites of pSVLGS.1) inserted at the BamHI site of pSV2.dhfr such that the GS and DHFR genes are in opposite orientations; 3) pSV2dhfr.GS6, which is identical to pSV2dhfr.GS3 except that the DHFR and GS genes are in the same orientation; and 4) pSV2dhfr.neI3, which contains a gene which confers resistance to the antibiotic G418 inserted at the BamHI site of pSV2.dhfr such that the two genes are in opposite orientations. 9cm petri dishes containing at least 10^6 cells were transfected with (i) 5ug or (ii) 10ug of each vector by calcium phosphate co-precipitation and the cells were allowed to recover in a non-selective medium. After 2 days, the medium was replaced by DMEM medium containing 10% dialysed FCS and 150ug/ml proline, to select for $dhfr^+$ transformants, or G418 in Ham's F12 medium to select for expression of the ne gene. To some dishes, methotrexate was also added to serve as an assay for the amount of DHFR enzyme produced. 9 days after transfection, the number of colonies on each plate was scored.

The results from two independent transfections are shown in Table 4. For transfection (i), 5ug of each plasmid was introduced into 10^6 cells on each dish. In transfection (ii) 10ug of DNA was used and the number of cells per dish was 4×10^6.

The concentrations of plasmid DNAs were carefully measured by absorbance (A_{260}) prior to transfection.

TABLE 4

Plasmid	no of surviving colonies/10^6 cells				
		MTX			
	OnM	50nM	(%*)	100nM	(%)
pSV2dhfr (i)	61	-	-	1	(1.6)
(ii)	240	13	(1.3)	1	
pSV2dhfrGS3 (i)	65	-		0	(-)
(ii)	1400	58	(4)	-	
pSV2dhfrGS6 (i)	90	-		5	(5.5)
(ii)	1800	90	(5)	-	
pSV2dhfrnel3 (i)	26	-		0	(-)
(ii)	152	0	(-)	-	

	G418 (0.8mg/ml)
pSV2dhfrnel3 (i)	77

*%: the percentage of the total dhfr$^+$ transfectants which are resistant to a given level of MTX.

The results show that the presence on the vector of a GS-minigene leads to the survival of a greater number of dhfr$^+$ transfectants and a greater proportion of these are resistant to high levels of MTX (50nM or 100nM) than if a vector lacking the GS-minigene is used.

This effect is observed only when the GS-minigene is in the same orientation in the vector as the DHFR gene, probably because convergent transcription when the genes are in opposite orientations is inhibitory for mRNA synthesis. This can also explain why the introduction of an irrelevant gene, the ne gene in pSV2dhfrnel3, in the opposite orientation leads to the survival of fewer dhfr$^+$ colonies than when pSV2dhfr is used for transfection.

The stimulating effect is not observed when a ne gene is inserted into pSV2dhfr, indicating that the enhanced expression of DHFR is a specific effect of a sequence(s) in the GS-minigene. Since an equal weight of each vector DNA was used for transfection and because pSV2dhfrGS6 is approximately twice the size of pSV2dhfr the stimulating effect is in fact observed even when the number of introduced molecules is only about half the number of pSV2dhfr molecules.

It can thus be seen that the use of the SHCS from the GS minigene leads to the production of transfectants having surprisingly and unexpectedly high copy number. The advantages of the use of such SHCSs in recombinant DNA technology are readily apparent.

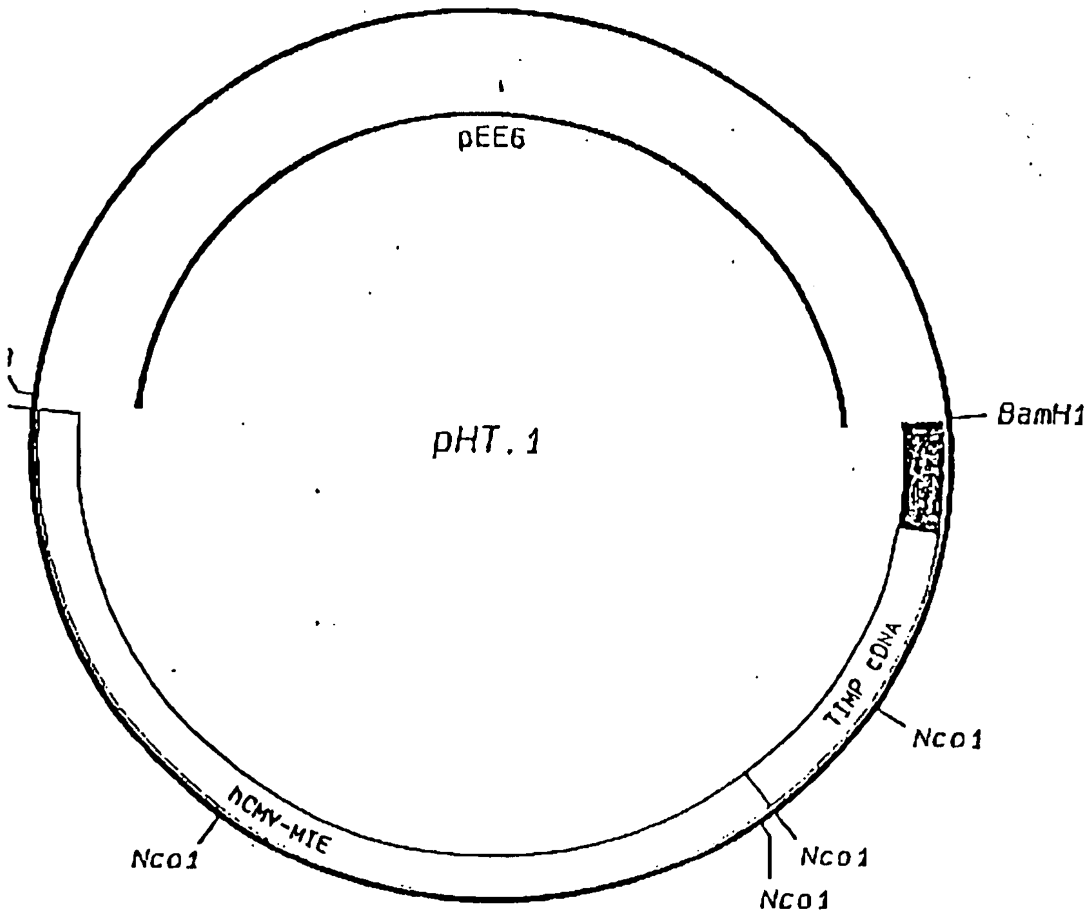


Figure 3

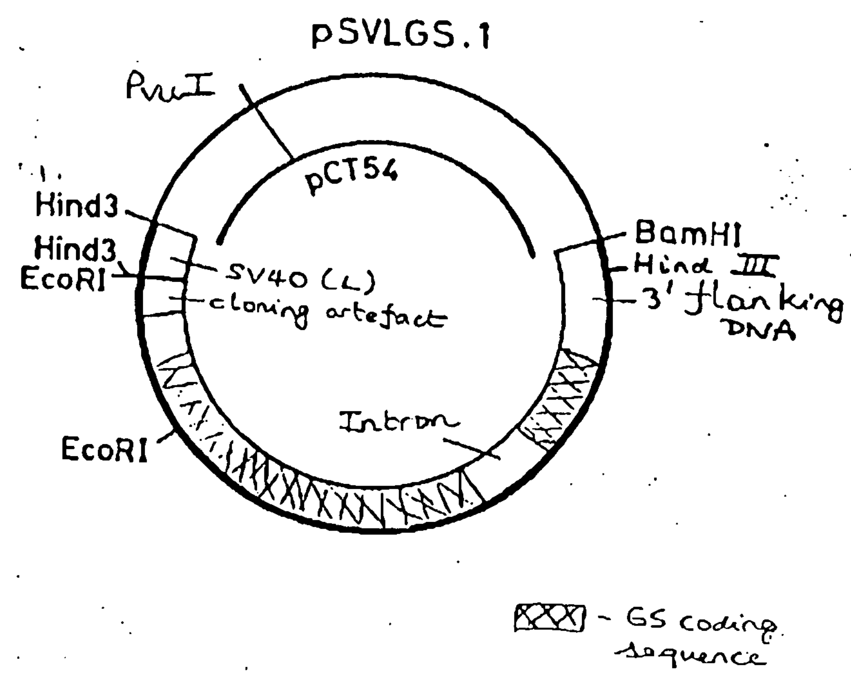


Figure 1

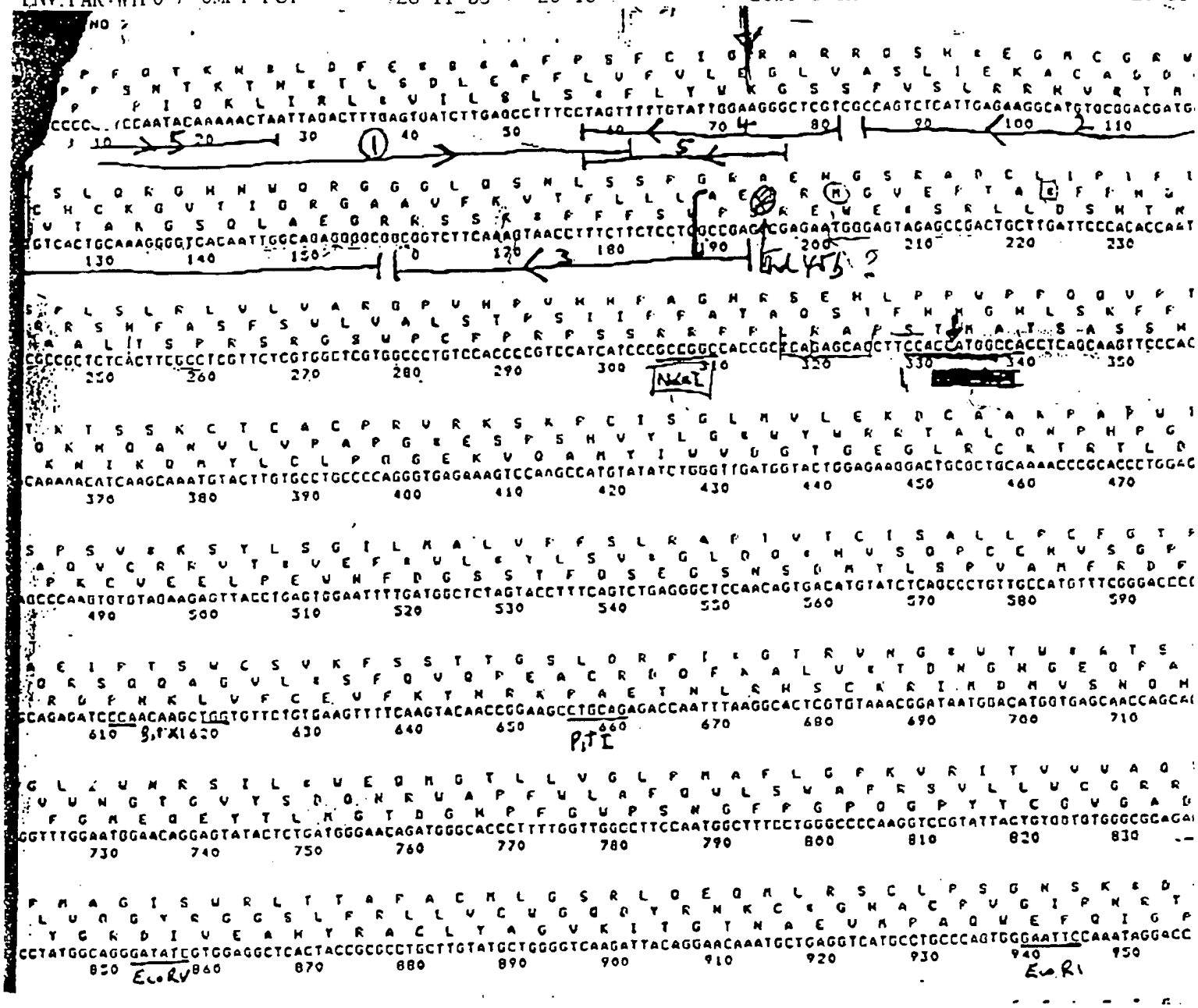


FIGURE 2A

FIGURE 2.2

GTGTC AAGGAATCGG CATGGGAGAT
CACAG TTCCTTAGCC GTACCCTCTA

a93 ← | → a90

7870 7880 7890 7900 7910 7920
G CTCTGGG TGGGGGGTTT CATGTTGGAT CGAGTATGTG AAGAGTTTGG GGTAATAGGA
CTACAGACCC ACCCCCCAAA GTACAACCTA GCTCATACAC TTCTCAAACC CCATTATCCT

7930 7940 7950 7960 7970 7980
ACCTTTGACC CCAAGCCGAT TCCTGGGAAC TGGAATGGTG CAGGCTGCCA TACGAAGTTT
TGGAAACTGG GGTTCGGCTA AGGACCCTTG ACCTTACCAC GTCCGACGGT ATGCTTCAAA

7990 8000 8010 8020 8030 8040
AGCACCAAGG CCATGCCGGA GGAGAATGGT CTGAAGTAAG TAGGCTGCTA TGGACCATCT
TCGTGGTTCC GGTACGGCCT CCTCTTACCA GACTTCATTC ATCCGACGAT ACCTGGTAGA

8050 8060 8070 8080 8090 8100
TTGTTCTC-- ---------- ---------- ---------- ---------- ----------
AACAAGAG-- ---------- ---------- ---------- ---------- ----------

8110 8120 8130 8140 8150 8160
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ----------

8170 8180 8190 8200 8210 8220
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ----------

8230 8240 8250 8260 8270 8280
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ----------

8290 8300 8310 8320 8330 PvuII 8340
---------- ---------- ---------- ---------- --------CA GCTGTGTGAT
---------- ---------- ---------- ---------- --------GT CGACACACTA

8350 8360 8370 AhaIII 8380 8390 8400
ATGGGAAAGT TGTCCACCTA GGTTCAAGCA TTTAAAGGTC TTTAGTAAGA ACTAAATACA
TACCCTTTCA ACAGGTGGAT CCAAGTTCGT AAATTTCCAG AAATCATTCT TGATTTATGT

8410 8420 8430 8440 8450 8460
CATACAAGCA AGTGGGTGAC TTAATTCTTA CTGATGGGAA GAGGCCAGTG ATGGGGGTCT
GTATGTTCGT TCACCCACTG AATTAAGAAT GACTACCCTT CTCCGGTCAC TACCCCCAGA

8470 8480 8490 8500 8510 8520
TCCATCCAAA AGATAATTGG TATTACATGT TAGGGGAGAT GGTGTGAAGC ACTTAGCAGC
AGGTAGGTTT TCTATTAACC ATAATGTACA ATCCCCTCTA CCACACTTCG TGAATCGTCG

8530 8540 8550 8560 8570 8580
AGATAGTACA GGGAGAGACA GGTCGGACTA AGGTATTTAT TGGTTTGTTA TAAGTCATCG
TCTATCATGT CCCTCTCTGT CCAGCCTGAT TCCATAAATA ACCAAACAAT ATTCAGTAGC

FIGURE 2C

```
AGCCAGAGGC GCGA------ --PPPP---- ---------- ---------- -TTCTGACAG

      8650       8660       8670       8680       8690       8700
T  AGCATTG AGAGATAGGT CACAAGGCAG AGACAGCCTG CATCAATATT TATTGTTTCT
ACTTCCTAAC TCTCTATCCA GTGTTCCGTC TCTGTCGGAC GTAGTTATAA ATAACAAAGA

      8710       8720       8730       8740       8750       8760
TGAACTCATG CCTCCCTCCT CCCCCTTGAA GGACAGGTTT CCTAGGTGAC AAGGTCAGAC
ACTTGAGTAC GGAGGGAGGA GGGGGAACTT CCTGTCCAAA GGATCCACTG TTCCAGTCTG

      8770       8780       8790       8800       8810       8820
CCTCACCTTT ACTGCTTCCA CCAGCCAGGC AGATGGAGGA GGGCATGGAG AAACTAAGCA
GGAGTGGAAA TGACGAAGGT GGTCGGTCCG TCTACCTCCT CCCGTACCTC TTTGATTCGT

      8830  KpnI 8840       8850       8860       8870       8880
AGGGGCACCG GTACCACATT CGAGGCTACG ATCCCAAGGG GGGCCTGGAC AATGCCCGTC
TCCCCGTGGC CATGGTGTAA GCTCCGATGC TAGGGTTCCC CCCGGACCTG TTACGGGCAG
   C10 <--|--> C7

      8890       8900       8910       8920       8930       8940
GTGTGACTGG GTTCCACGAA ACCTCCAACA TGAAGGCATT TTCTGCTGGT GTCGCCAAGT
CACACTGACC CAAGGTGCTT TGGAGGTTGT ACTTCCGTAA AAGACGACCA CAGCGGTTCA

      8950       8960       8970       8980       8990       9000
CGAATCGATC CGATTGCGAG AGAATTATTA AGACGCCCCC TCTGCAATGT GACCTTGCAG
GCTTAGCTAG GCTAACGCTC TCTTAATAAT TCTGCGGGGG AGACGTTACA CTGGAACGTC

      9010       9020       9030       9040       9050       9060
TGACAGAAGC CATCGTGGGC ACATGCCTTC TGAATGAGAG TCCCGACGGA GCCCTTCGAA
ACTGTCTTCG GTAGCACCCG TGTACGGAAG ACTTACTCTC AGGGCTGCCT CGGGAAGCTT

      9070       9080       9090       9100       9110       9120
TA-------- ---------- ---------- -------CAA AAACTAATTA GACTTTGAGT
AT-------- ---------- ---------- -------GTT TTTGATTAAT CTGAAACTCA

      9130       9140       9150  PvuII 9160      9170       9180
GATCTTGAGC CTTTCCTAGT TCATGGCAGC CCGCCCCAGC TGTGTCATTG TAACTCAAAG
CTAGAACTCG GAAAGGATCA AGTACCGTCG GGCGGGGTCG ACACAGTAAC ATTGAGTTTC
                                 C4 <--|--> C72

      9190       9200       9210       9220       9230       9240
GATGGAATAT CAACGGTCTT TTTATTCCTC GTGCCCAGTT AATCCTTGCT TTTATTGGTC
CTACCTTATA GTTGCCAGAA AAATAAGGAG CACGGGTCAA TTAGGAACGA AAATAACCAG

      9250       9260       9270       9280       9290       9300
AGAATAGAGG AGTCAAGTTC TTAATGCCTA TACACCAACG TCATTTGTTT TCTATTTAGC
TCTTATCTCC TCAGTTCAAG AATTACGGAT ATGTGGTTGC AGTAAACAAA AGATAAATCG

      9310       9320       9330       9340       9350       9360
TTTCTACCTC GGGGTGGGAG GGGTAGGGAG GGGTAGGCCA AGCCAACCTA ACCACATGCT
AAAGATGGAG CCCCACCCTC CCCATCCCTC CCCATCCGGT TCGGTTGGAT TGGTGTACGA
```

– FIGURE 2b

```
       9430       9440       9450       9460       9470       9480
TTT ACG--- ---------- ---------- ---------- ---------C AGACAGCCTA
AA TGC--- ---------- ---------- ---------- ---------C TCTGTCGGAT

       9490       9500       9510       9520       9530       9540
TTTACACCGT ACAGATCCTA CCCTACGTTA AAGAGAAAGT TATTTCTGGT CCTCCAGTTA
AAATGTGGCA TGTCTAGGAT GGGATGCAAT TTCTCTTTCA ATAAAGACCA GGAGGTCAAT

                      AhaIII
       9550       9560       9570       9580       9590       9600
TAACACAAGC AGATGATATT TTATATTTAA ATGTAAAAAC AAAAGTTATA TATATGATAT
ATTGTGTTCG TCTACTATAA AATATAAATT TACATTTTTG TTTTCAATAT ATATACTATA

       9610       9620       9630       9640       9650       9660
GTGGATATAT GTGTATTTCT AATTCAGAAA CCATCCTAGT TACTGGGTTT GCCAAGTTTG
CACCTATATA CACATAAAGA TTAAGTCTTT GGTAGGATCA ATGACCCAAA CGGTTCAAAC

       9670 HpaI 9680       9690       9700       9710       9720 Bst
AACAGCTTGG TTAACAGAAA AGGATCTCTT GAGTAGAGGT GGGGGTGCAG TACGAGGAAA
TTGTCGAACC AATTGTTCTT TCCTAGAGAA CTCATCTCCA CCCCCACGTC ATGCTCCTTT
c75 ←

                                                      PstI
       9730       9740       9750       9760       9770       9780
G  GTTATCT GGGGCTCAGC GCTTTATTAC TATGTGGGGT TTCCCTGCCC ACTCTGCAGG
CACCAATAGA CCCCGAGTCG CGAAATAATG ATACACCCCA AAGGGACGGG TGAGACGTCC
                                               c15     ← → c101
                                               c269

       9790       9800       9810       9820       9830       9840
AGCAGATGCT GGACAGGTAG CCAGGGTGGG AGAGAGTGGT TGCCACCACC TGTGCCTGTG
TCGTCTACGA CCTGTCCATC GGTCCCACCC TCTCTCACCA ACGGTGGTGG ACACGGACAC

                                                      HincII
       9850       9860       9870       9880       9890       9900
CTTAGGCCTA AGATGCATAT GTATCCACAC AGAGTTAGGA GGATGGAGTT GGCTGGTCAA
GAATCCGGAT TCTACGTATA CATAGGTGTG TCTCAATCCT CCTACCTCAA CCGACCAGTT

       9910       9920       9930       9940       9950       9960
CTTGAAGATT TGTTACTGAT AGGGGTGGTG GGTTTATTTT TTGGTGGCAT AGCATGTCAC
GAACTTCTAA ACAATGACTA TCCCCACCAC CCAAATAAAA AACCACCGTA TCGTACAGTG

                                  AhaIII
       9970       9980       9990      10000      10010      10020
ATAAAGCAGG GCTTTGATAT ATTAAATTTT TTTAAAGCAA ACATGTTCAG CTTTATCACC
TATTTCGTCC CGAAACTATA TAATTTAAAA AAATTTCGTT TGTACAAGTC GAAATAGTGG

      10030      10040      10050      10060      10070      10080
TTGTAGGGTT TCTAGTT--- ---------- ---------- ---------- ----------
AACATCCCAA AGATCAA--- ---------- ---------- ---------- ----------

      10090      10100      10110      10120      10130      10140
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ----------
```

```
      10310       10320       10330       10340       10350       10360
--- ------ ACC1------ ---------- ---------- ---------- ---------- ----------
- -------- TGG4------ ---------- ---------- --  ------- ----------

      10270       10280       10290       10300       10310       10320
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ----------

      10330       10340       10350       10360       10370       10380
---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ----------

      10390       10400       10410       10420       10430       10440
---------- ------- -- ---------- --------- ---------- ----------
---------- ---------- ------ --- ---------- ---------- ----------

      10450       10460       10470       10480       10490       10500
 --------- -----CCATG TACGATATGT TCGATTACAT AACTCCTCAT GTAAACACAC
---------- -----GGTAC ATGCTATACA AGCTAATGTA TTGAGGAGTA CATTTGTGTG

      10510       10520       10530       10540       10550       10560
T  AACTGCC ACACCAGCGC TTATAAATCA ACCTAACATT TATAAGATTT CCTCTTGACT
ALATTGACGG TGTCGTCGCG AATATTTAGT TGGATTGTAA ATATTCTAAA GGAGAACTGA

      10570       10580       10590       10600       10610       10620
TGTTTCTTTC TGCTTGGCGC ACGAACGAAA AAAAAACTGC GATATTTTTT TGTTCCTTCA
ACAAACGAAAC ACCAACCGCG TGCTTGTTTT TTTTTTGACG CTATAAAAAA ACAAGGAAGT

      10630       10640       10650       10660       10670
TTTCCTATCA AAAGAAACGC GACTGGTTCT GTTTTGTTCA CCCGCAAAAT AAGCTT
AAAGGATAGT TTTCTTTCCG CTCACCAAGA CAAAACAAAT GACCGTTTTA TTCGAA
```

HindIII 4

4

C108

All finished

-LGT

FIGURE 2E

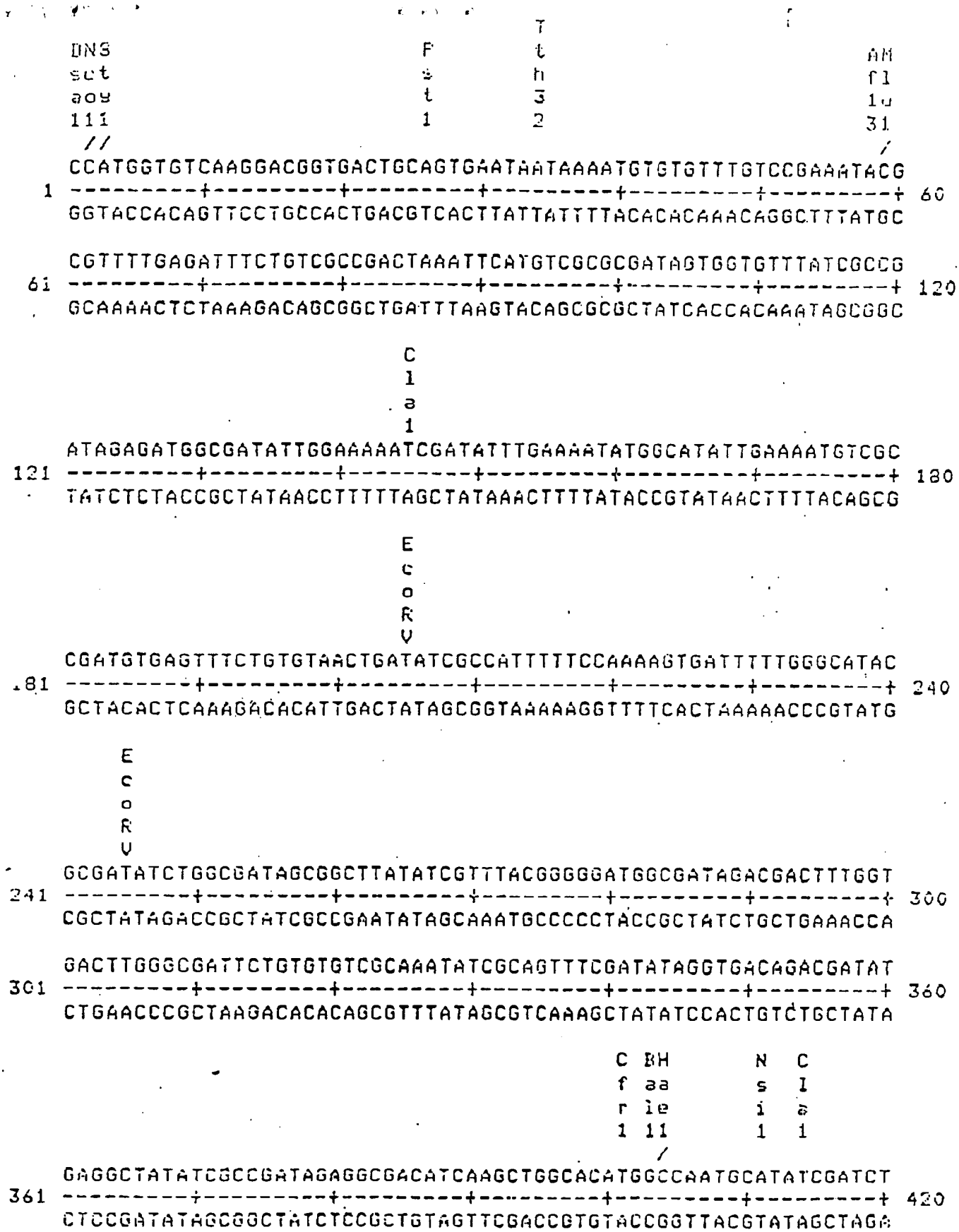

DNS sct aoB 111
F s t 1
T t h 3 2
AM f1 1u 31
CCATGGTGTCAAGGACGGTGACTGCAGTGAATAATAAAATGTGTGTTTGTCCGAAATACG
GGTACCACAGTTCCTGCCACTGACGTCACTTATTATTTTACACACAAACAGGCTTTATGC
1
60
CGTTTTGAGATTTCTGTCGCCGACTAAATTCATGTCGCGCGATAGTGGTGTTTATCGCCG
GCAAAACTCTAAAGACAGCGGCTGATTTAAGTACAGCGCGCTATCACCACAAATAGCGGC
61
120
C l a 1
ATAGAGATGGCGATATTGGAAAAATCGATATTTGAAAATATGGCATATTGAAAATGTCGC
TATCTCTACCGCTATAACCTTTTTAGCTATAAACTTTTATACCGTATAACTTTTACAGCG
121
180
E c o R V
CGATGTGAGTTTCTGTGTAACTGATATCGCCATTTTTCCAAAAGTGATTTTTGGGCATAC
GCTACACTCAAAGACACATTGACTATAGCGGTAAAAAGGTTTTCACTAAAAACCCGTATG
181
240
E c o R V
GCGATATCTGGCGATAGCGGCTTATATCGTTTACGGGGGATGGCGATAGACGACTTTGGT
CGCTATAGACCGCTATCGCCGAATATAGCAAATGCCCCCTACCGCTATCTGCTGAAACCA
241
300
GACTTGGGCGATTCTGTGTGTCGCAAATATCGCAGTTTCGATATAGGTGACAGACGATAT
CTGAACCCGCTAAGACACACAGCGTTTATAGCGTCAAAGCTATATCCACTGTCTGCTATA
301
360
C B H f aa r le 1 11
N C s I i a 1 1
GAGGCTATATCGCCGATAGAGGCGACATCAAGCTGGCACATGGCCAATGCATATCGATCT
CTCCGATATAGCGGCTATCTCCGCTGTAGTTCGACCGTGTACCGGTTACGTATAGCTAGA
361
420

Figure 4A


S C BH
s f aa
p r le
1 1 11
ATACATTGAATCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCA
421
TATGTAACTTAGTTATAACCGGTAATCGGTATAATAAGTAACCAATATATCGTATTTAGT
480
S C BH
s f aa
p r le
1 1 11
ATATTGGCTATTGGCCATTGCATACGTTGTATCCATATCATAATATGTACATTTATATTG
481
TATAACCGATAACCGGTAACGTATGCAACATAGGTATAGTATTATACATGTAAATATAAC
540
H
i M S
n m p
c e e
2 1 1
GCTCATGTCCAACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAAT
541
CGAGTACAGGTTGTAATGGCGGTACAACTGTAACTAATAACTGATCAATAATTATCATTA
600
CAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGG
601
GTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCC
660
B A A
g h a
l a t
1 2 2
TAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGT
661
ATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCA
720
A A
h a
a t
2 2
ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC
721
TACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATG
780
B N
g d
l e
1 1
GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTG
781
CCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAAC
840

FIGURE 4B

```
      A  A                    B
      h  a                    s
      a  t                    1
      2  2                    1
     ACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACT
 541 ---------+---------+---------+---------+---------+---------+ 900
     TGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGA

                              S
                              n                    DNS
                              a                    sct
                              B                    aoy
                              1                    111
                                                   //
     TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT
 901 ---------+---------+---------+---------+---------+---------+ 960
     AAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAA

     GGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC
 961 ---------+---------+---------+---------+---------+---------+ 1020
     CCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGG

           A  A                    B
           h  a                    a
           a  t                    n
           2  2                    1
     CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC
  21 ---------+---------+---------+---------+---------+---------+ 1080
     GGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAG

     GTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA
1081 ---------+---------+---------+---------+---------+---------+ 1140
     CATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATAT

              BH
             BsgS                    G   A
             aPia                    s   h
             n1Ac                    u   a
             2211                    1   2
              ///
     TAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTG
1141 ---------+---------+---------+---------+---------+---------+ 1200
     ATTCGTCTCGAGCAAATCACTTGGCAGTCTAGCGGACCTCTGCGGTAGGTGCGACAAAAC

                                      N
                 B                D BCGsSX
                 b                s sfdPam
                 v                a lriBca
                 2                1 112223
                                    ////
     ACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGGGAACGGTGCATTGGAA
1201 ---------+---------+---------+---------+---------+---------+ 1260
     TGGAGGTATCTTCTGTGGCCCTGGCTAGGTCGGAGGCGCCGGCCCTTGCCACGTAACCTT
```

Figure 4C

```
      CGCGGATTCCCCGTGCCAAGAGTGACGTAAGTACCGCCTATAGAGTCTATAGGCCCACCC
1271  ---------+---------+---------+---------+---------+---------+ 1320
      GCGCCTAAGGGGCACGGTTCTCACTGCATTCATGGCGGATATCTCAGATATCCGGGTGGG

        B                 N
      Ss                N sS
      tt                s pp
      gX                i Hn
      11                1 11
                          /
      CCTTGGCTTCTTATGCATGCTATACTGTTTTTGGCTTGGGGTCTATACACCCCCGCTTCC
1321  ---------+---------+---------+---------+---------+---------+ 1380
      GGAACCGAAGAATACGTACGATATGACAAAAACCGAACCCCAGATATGTGGGGGCGAAGG

                               E
                               s
                               p
                               1
      TCATGTTATAGGTGATGGTATAGCTTAGCCTATAGGTGTGGGTTATTGACCATTATTGAC
1381  ---------+---------+---------+---------+---------+---------+ 1440
      AGTACAATATCCACTACCATATCGAATCGGATATCCACACCCAATAACTGGTAATAACTG

                                                    P
                                                    f
                                                    1
                                                    M
                                                    1
      CACTCCCCTATTGGTGACGATACTTTCCATTACTAATCCATAACATGGCTCTTTGCCACA
1441  ---------+---------+---------+---------+---------+---------+ 1500
      GTGAGGGGATAACCACTGCTATGAAAGGTAATGATTAGGTATTGTACCGAGAAACGGTGT

                          E
                          c
                          o
                          5
                          7
      ACTCTCTTTATTGGCTATATGCCAATACACTGTCCTTCAGAGACTGACACGGACTCTGTA
1501  ---------+---------+---------+---------+---------+---------+ 1560
      TGAGAGAAATAACCGATATACGGTTATGTGACAGGAAGTCTCTGACTGTGCCTGAGACAT

                          E
                          c
                          o
                          3
                          1
      TTTTTACAGGATGGGGTCTCATTTATTATTTACAAATTCACATATACAACACCACCGTCC
1561  ---------+---------+---------+---------+---------+---------+ 1620
      AAAAATGTCCTACCCCAGAGTAAATAATAAATGTTTAAGTGTATATGTTGTGGTGGCAGG

                B
                s                         X               A       A
                p                         h               v       f
                1                         o               a       1
                2                         2               i       3
      CCAGTGCCCGCAGTTTTTATTAAACATAACGTGGGATCTCCACGCGAATCTCGGGTACGT
1621  ---------+---------+---------+---------+---------+---------+ 1680
      GGTCACGGGCGTCAAAAATAATTTGTATTGCACCCTAGAGGTGCGCTTAGAGCCCATGCA
```

FIGURE 4D

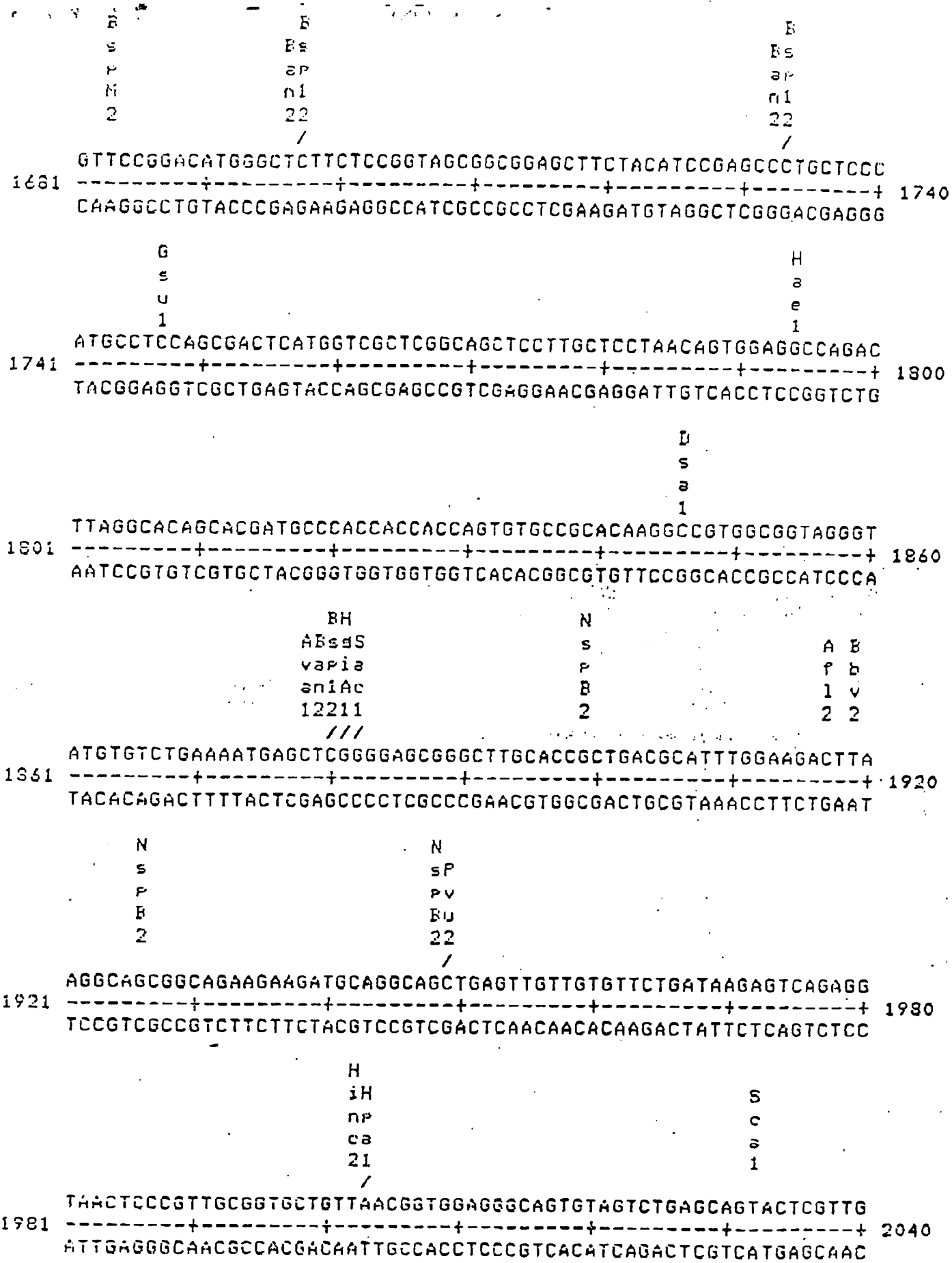

BspMI22
BspMI22
BspMI22
GTTCCGGACATGGGCTCTTCTCCGGTAGCGGCGGAGCTTCTACATCCGAGCCCTGCTCCC
1681
CAAGGCCTGTACCCGAGAAGAGGCCATCGCCGCCTCGAAGATGTAGGCTCGGGACGAGGG
1740
GseI1
HaeI1
ATGCCTCCAGCGACTCATGGTCGCTCGGCAGCTCCTTGCTCCTAACAGTGGAGGCCAGAC
1741
TACGGAGGTCGCTGAGTACCAGCGAGCCGTCGAGGAACGAGGATTGTCACCTCCGGTCTG
1800
DsaI1
TTAGGCACAGCACGATGCCCACCACCACCAGTGTGCCGCACAAGGCCGTGGCGGTAGGGT
1801
AATCCGTGTCGTGCTACGGGTGGTGGTGGTCACACGGCGTGTTCCGGCACCGCCATCCCA
1860
BHABssSvaPIaanIAc12211
NsPB2
ABfbIv22
ATGTGTCTGAAAATGAGCTCGGGGAGCGGGCTTGCACCGCTGACGCATTTGGAAGACTTA
1861
TACACAGACTTTTACTCGAGCCCCTCGCCCGAACGTGGCGACTGCGTAAACCTTCTGAAT
1920
NsPB2
NsPvBu22
AGGCAGCGGCAGAAGAAGATGCAGGCAGCTGAGTTGTTGTGTTCTGATAAGAGTCAGAGG
1921
TCCGTCGCCGTCTTCTTCTACGTCCGTCGACTCAACAACACAAGACTATTCTCAGTCTCC
1980
HiHnPca21
ScaI1
TAACTCCCGTTGCGGTGCTGTTAACGGTGGAGGGCAGTGTAGTCTGAGCAGTACTCGTTG
1981
ATTGAGGGCAACGCCACGACAATTGCCACCTCCCGTCACATCAGACTCGTCATGAGCAAC
2040

Figure 4E

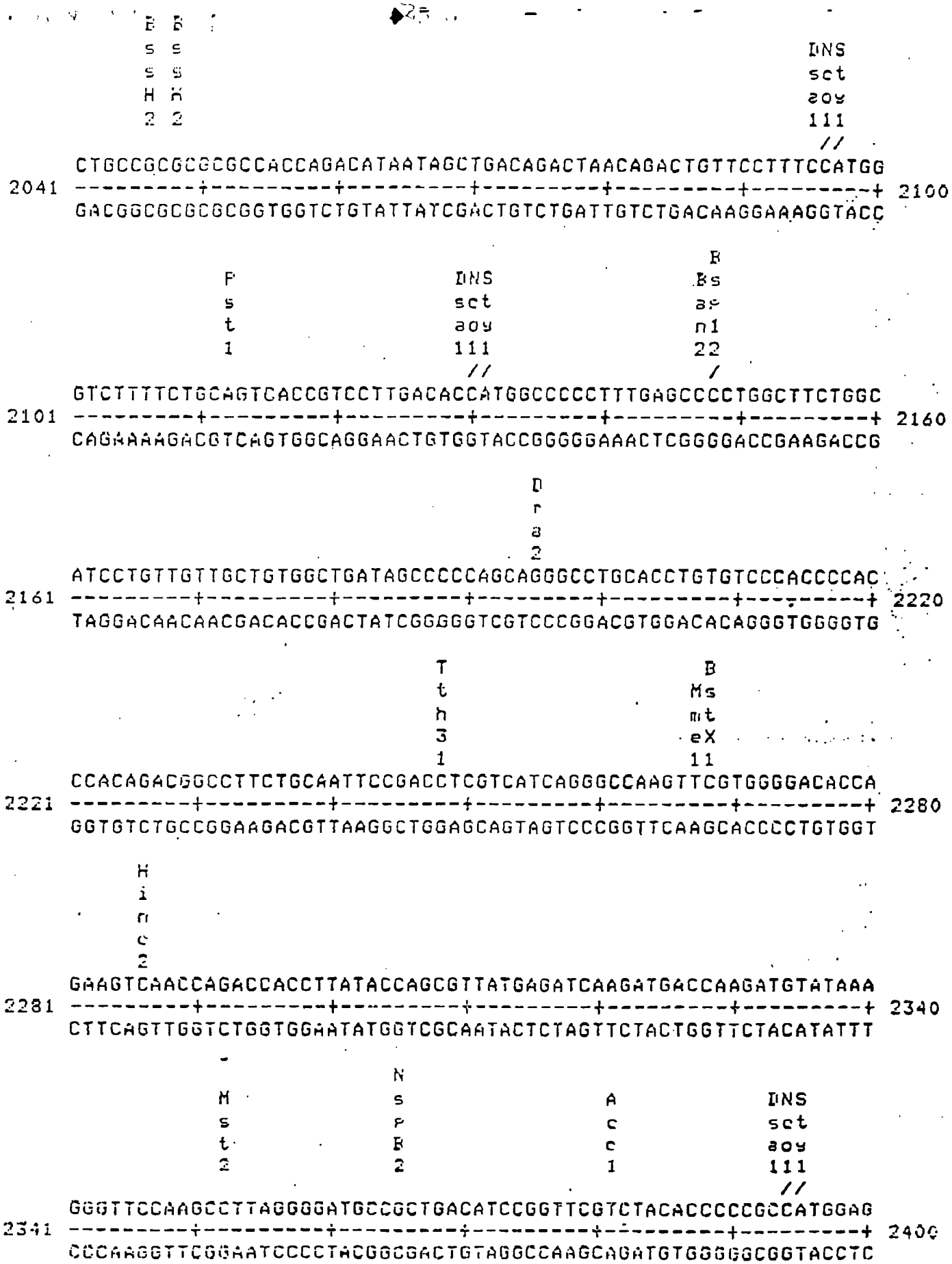

B B
s s
s s
H H
2 2
DNS
sct
aoy
111
//
CTGCCGCGCGCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCCTTTCCATGG
2041 ---------+---------+---------+---------+---------+---------+ 2100
GACGGCGCGCGCGGTGGTCTGTATTATCGACTGTCTGATTGTCTGACAAGGAAAGGTACC
P
s
t
1
DNS
sct
aoy
111
//
B
Bs
ap
n1
22
/
GTCTTTTCTGCAGTCACCGTCCTTGACACCATGGCCCCCTTTGAGCCCCTGGCTTCTGGC
2101 ---------+---------+---------+---------+---------+---------+ 2160
CAGAAAAGACGTCAGTGGCAGGAACTGTGGTACCGGGGGAAACTCGGGGACCGAAGACCG
D
r
a
2
ATCCTGTTGTTGCTGTGGCTGATAGCCCCCAGCAGGGCCTGCACCTGTGTCCCACCCCAC
2161 ---------+---------+---------+---------+---------+---------+ 2220
TAGGACAACAACGACACCGACTATCGGGGGTCGTCCCGGACGTGGACACAGGGTGGGGTG
T
t
h
3
1
B
Ms
mt
eX
11
CCACAGACGGCCTTCTGCAATTCCGACCTCGTCATCAGGGCCAAGTTCGTGGGGACACCA
2221 ---------+---------+---------+---------+---------+---------+ 2280
GGTGTCTGCCGGAAGACGTTAAGGCTGGAGCAGTAGTCCCGGTTCAAGCACCCCTGTGGT
H
i
n
c
2
GAAGTCAACCAGACCACCTTATACCAGCGTTATGAGATCAAGATGACCAAGATGTATAAA
2281 ---------+---------+---------+---------+---------+---------+ 2340
CTTCAGTTGGTCTGGTGGAATATGGTCGCAATACTCTAGTTCTACTGGTTCTACATATTT
-
M
s
t
2
N
s
p
B
2
A
c
c
1
DNS
sct
aoy
111
//
GGGTTCCAAGCCTTAGGGGATGCCGCTGACATCCGGTTCGTCTACACCCCCGCCATGGAG
2341 ---------+---------+---------+---------+---------+---------+ 2400
CCCAAGGTTCGGAATCCCCTACGGCGACTGTAGGCCAAGCAGATGTGGGGGCGGTACCTC

FIGURE 4F

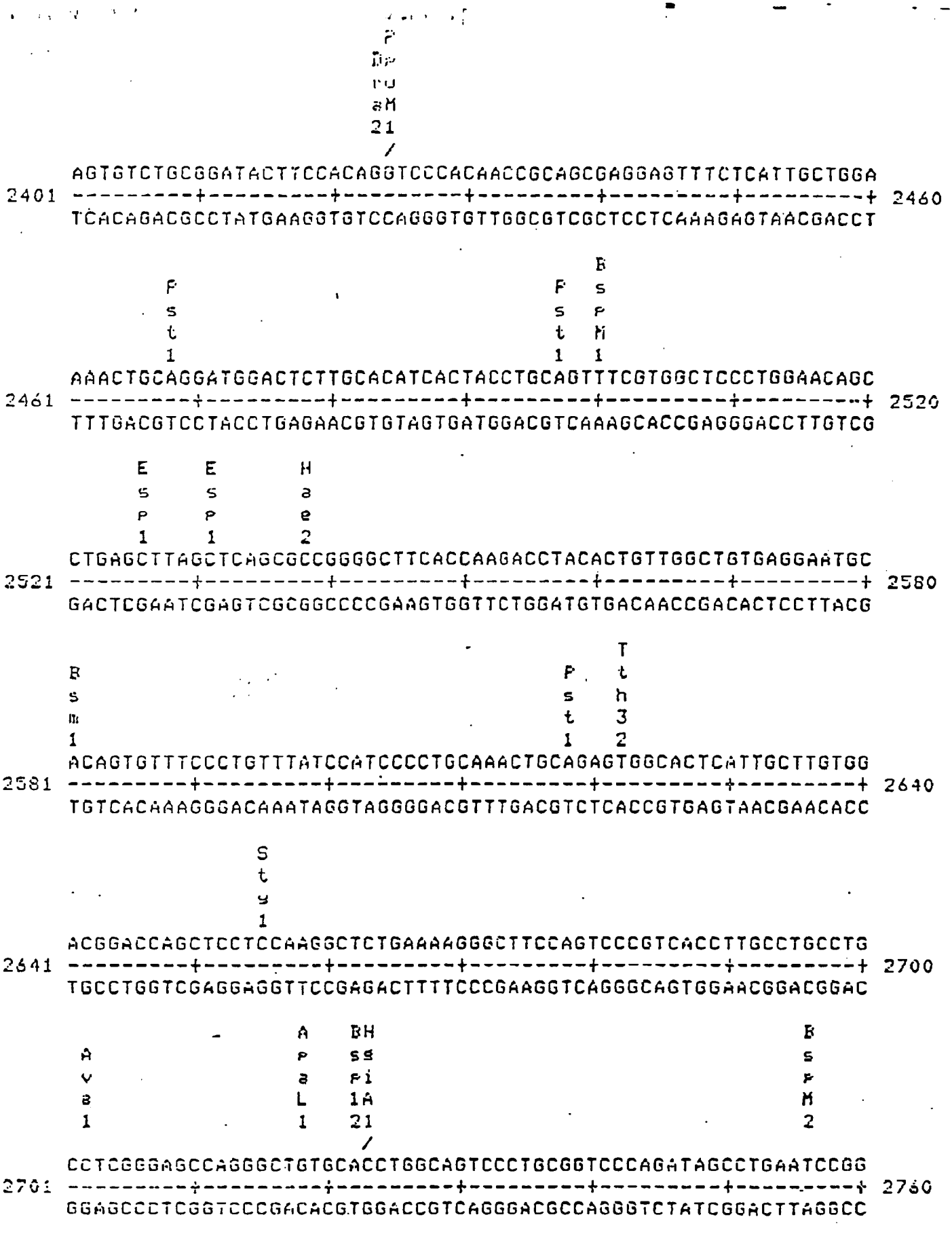

P
Dp
nI
aM
21
AGTGTCTGCGGATACTTCCACAGGTCCCACAACCGCAGCGAGGAGTTTCTCATTGCTGGA
2401 ---------+---------+---------+---------+---------+---------+ 2460
TCACAGACGCCTATGAAGGTGTCCAGGGTGTTGGCGTCGCTCCTCAAAGAGTAACGACCT
B
P s
s P
t M
1 1
AAACTGCAGGATGGACTCTTGCACATCACTACCTGCAGTTTCGTGGCTCCCTGGAACAGC
2461 ---------+---------+---------+---------+---------+---------+ 2520
TTTGACGTCCTACCTGAGAACGTGTAGTGATGGACGTCAAAGCACCGAGGGACCTTGTCG
E E H
s s a
P P e
1 1 2
CTGAGCTTAGCTCAGCGCCGGGGCTTCACCAAGACCTACACTGTTGGCTGTGAGGAATGC
2521 ---------+---------+---------+---------+---------+---------+ 2580
GACTCGAATCGAGTCGCGGCCCCGAAGTGGTTCTGGATGTGACAACCGACACTCCTTACG
T
B P t
s s h
m t 3
1 1 2
ACAGTGTTTCCCTGTTTATCCATCCCCTGCAAACTGCAGAGTGGCACTCATTGCTTGTGG
2581 ---------+---------+---------+---------+---------+---------+ 2640
TGTCACAAAGGGACAAATAGGTAGGGGACGTTTGACGTCTCACCGTGAGTAACGAACACC
S
t
y
1
ACGGACCAGCTCCTCCAAGGCTCTGAAAAGGGCTTCCAGTCCCGTCACCTTGCCTGCCTG
2641 ---------+---------+---------+---------+---------+---------+ 2700
TGCCTGGTCGAGGAGGTTCCGAGACTTTTCCCGAAGGTCAGGGCAGTGGAACGGACGGAC
A BH B
A P s s
V a P i P
B L 1 A M
1 1 21 2
CCTCGGGAGCCAGGGCTGTGCACCTGGCAGTCCCTGCGGTCCCAGATAGCCTGAATCCGG
2701 ---------+---------+---------+---------+---------+---------+ 2760
GGAGCCCTCGGTCCCGACACGTGGACCGTCAGGGACGCCAGGGTCTATCGGACTTAGGCC

FIGURE 4G

```
                                                                    D
                                                                    r
                                                                    a
                                                                    1
       ATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACAC
  2761 ---------+---------+---------+---------+---------+---------+ 2820
       TAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTG

                                                                    H
                                                  B                 iH
                                                  s                 np
                                                  m                 ca
                                                  1                 21
                                                                     /
       CTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCA
  2821 ---------+---------+---------+---------+---------+---------+ 2880
       GAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGT

       GCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTT
  2881 ---------+---------+---------+---------+---------+---------+ 2940
       CGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAA

                                                                B
          B                                                     aX
          s                                                     mh
          m                                                     Ho
          1                                                     12
                                                                 /
       TCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATC
  2941 ---------+---------+---------+---------+---------+---------+ 3000
       AGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAG

       C
  3001 - 3001
       G
```

Enzymes that do cut:

```
Aat2   Acc1   Afl2   Afl3   Aha2   ApaL1  Ava1   Bal1   BamH1  Ban1   Ban2
Bbv2   Bgl1   Bsm1   Bsp12  BspM1  BspM2  BssH2  BstX1  Cfr1   Cla1   Dra1
Dra2   Dsa1   Eco31  Eco57  EcoRV  Esp1   Gdi2   Gsu1   Hae1   Hae2   HgiA1
Hinc2  Hpa1   Mlu1   Mme1   Mst2   Nco1   Nde1   Nsi1   NspB2  NspH1  PflM1
PpuM1  Pst1   Pvu2   Sac1   Sac2   Sca1   SnaB1  Spe1   Sph1   Ssp1   Sty1
Tth31  Tth32  Xho2   Xma3
```

Enzymes that do not cut:

```
Apa1   Asu2   Avr2   Bcl1   Bgl2   BspH1  BstE2  Cfr10  Dra3   EcoB   EcoK
EcoR1  Fsp1   HgiE2  Hind3  Kpn1   Nae1   Nar1   Nhe1   Not1   Nru1   PmaC1
Pvu1   Rsr2   Sal1   Sfi1   Sma1   Spl1   Stu1   Xba1   Xho1   Xmn1
```

FIGURE 4H

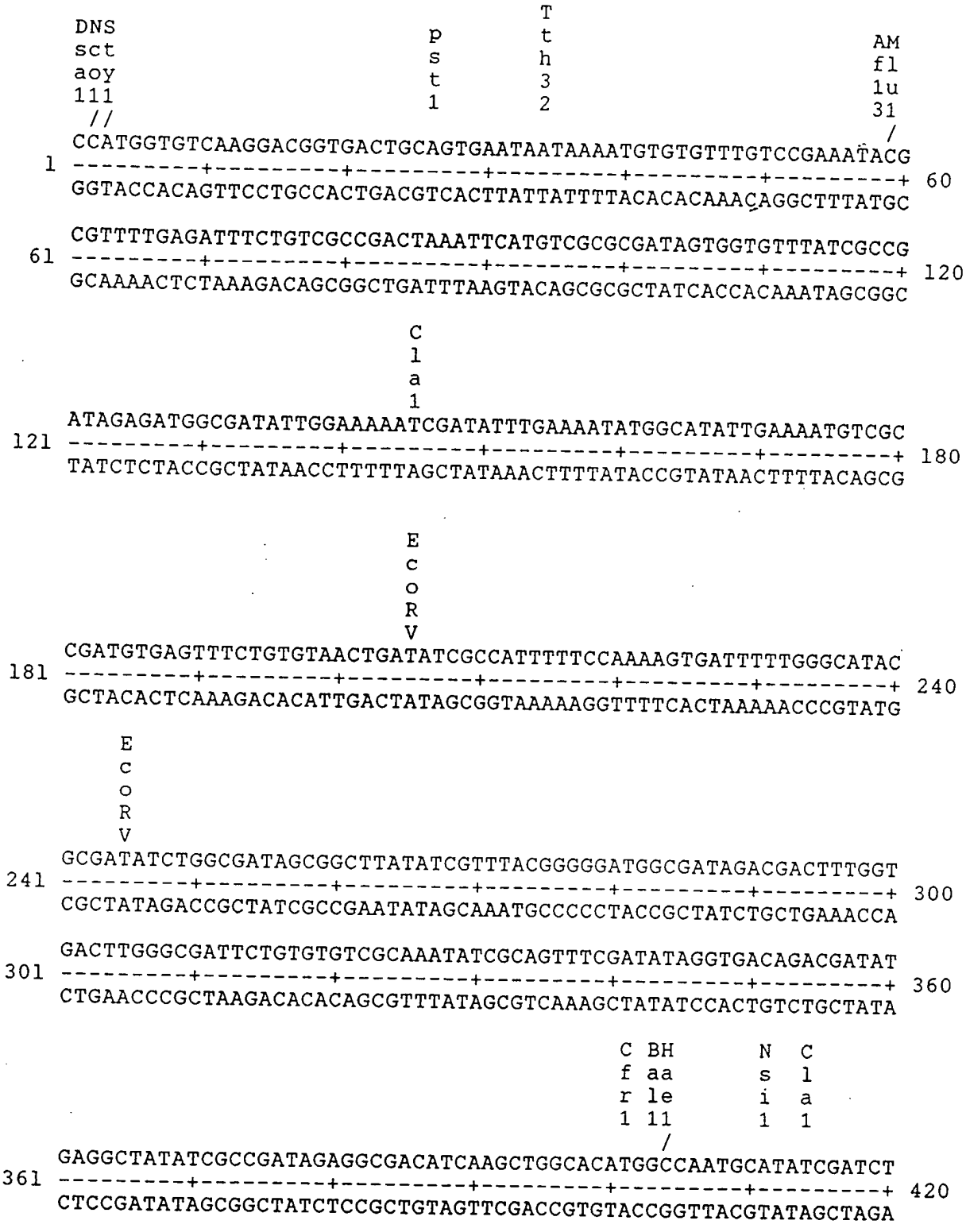

DNS
sct
aoy
111
Pst1
Tth32
AMfl1u31
CCATGGTGTCAAGGACGGTGACTGCAGTGAATAATAAAATGTGTGTTTGTCCGAAATACG
GGTACCACAGTTCCTGCCACTGACGTCACTTATTATTTTACACACAAACAGGCTTTATGC
CGTTTTGAGATTTCTGTCGCCGACTAAATTCATGTCGCGCGATAGTGGTGTTTATCGCCG
GCAAAACTCTAAAGACAGCGGCTGATTTAAGTACAGCGCGCTATCACCACAAATAGCGGC
ClaI
ATAGAGATGGCGATATTGGAAAAATCGATATTTGAAAATATGGCATATTGAAAATGTCGC
TATCTCTACCGCTATAACCTTTTTAGCTATAAACTTTTATACCGTATAACTTTTACAGCG
EcoRV
CGATGTGAGTTTCTGTGTAACTGATATCGCCATTTTTCCAAAAGTGATTTTTGGGCATAC
GCTACACTCAAAGACACATTGACTATAGCGGTAAAAAGGTTTTCACTAAAAACCCGTATG
EcoRV
GCGATATCTGGCGATAGCGGCTTATATCGTTTACGGGGGATGGCGATAGACGACTTTGGT
CGCTATAGACCGCTATCGCCGAATATAGCAAATGCCCCCTACCGCTATCTGCTGAAACCA
GACTTGGGCGATTCTGTGTGTCGCAAATATCGCAGTTTCGATATAGGTGACAGACGATAT
CTGAACCCGCTAAGACACACAGCGTTTATAGCGTCAAAGCTATATCCACTGTCTGCTATA
CfrI1 BHaIIe11 NsiI1 ClaI1
GAGGCTATATCGCCGATAGAGGCGACATCAAGCTGGCACATGGCCAATGCATATCGATCT
CTCCGATATAGCGGCTATCTCCGCTGTAGTTCGACCGTGTACCGGTTACGTATAGCTAGA

Fig. 4A

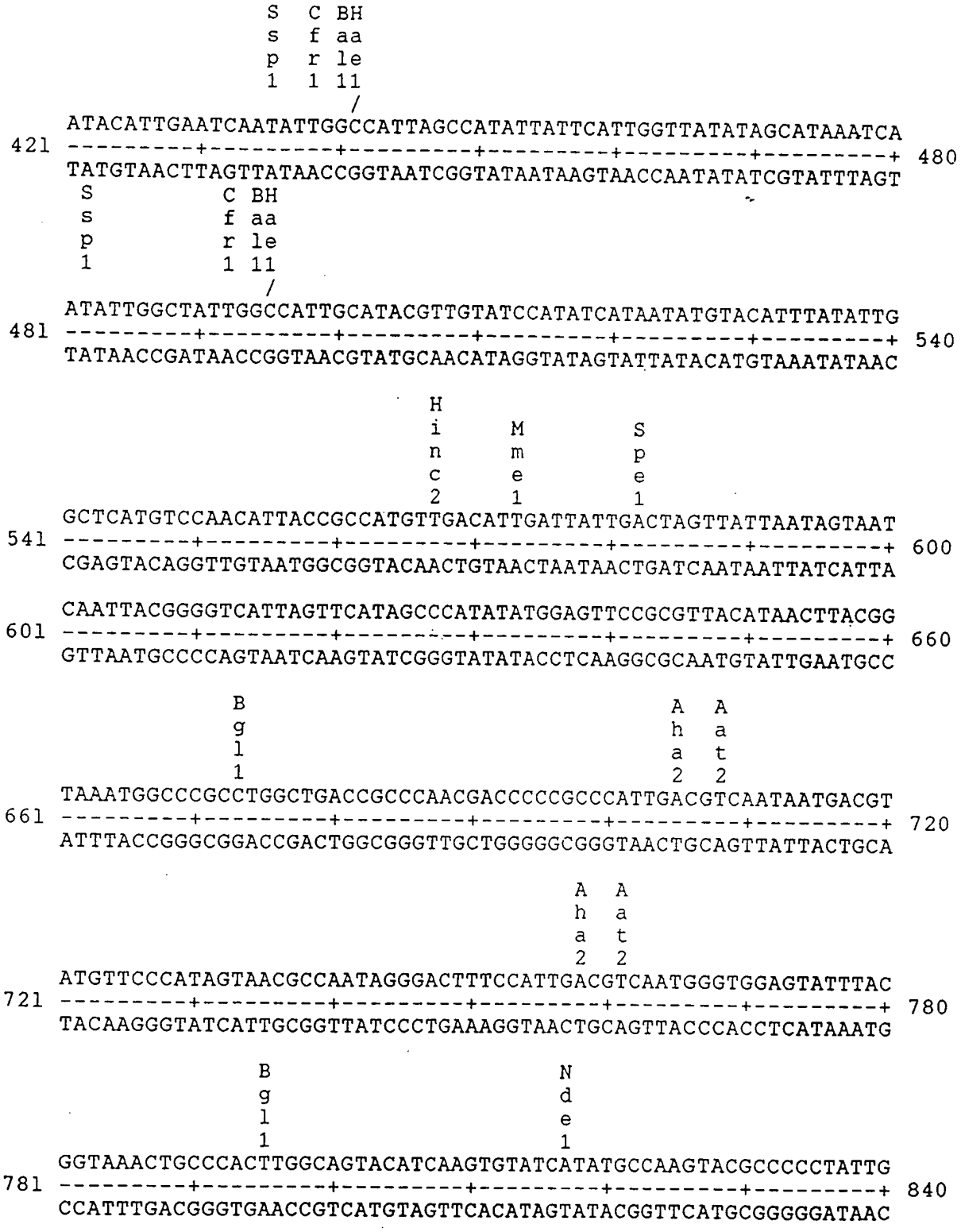

S C BH
s f aa
p r le
1 1 11
ATACATTGAATCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCA
421
TATGTAACTTAGTTATAACCGGTAATCGGTATAATAAGTAACCAATATATCGTATTTAGT
480
S C BH
s f aa
p r le
1 1 11
ATATTGGCTATTGGCCATTGCATACGTTGTATCCATATCATAATATGTACATTTATATTG
481
TATAACCGATAACCGGTAACGTATGCAACATAGGTATAGTATTATACATGTAAATATAAC
540
H M S
i m p
n e e
c 2 1 1
GCTCATGTCCAACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAAT
541
CGAGTACAGGTTGTAATGGCGGTACAACTGTAACTAATAACTGATCAATAATTATCATTA
600
CAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGG
601
GTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCC
660
B A A
g h a
l a t
1 2 2
TAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGT
661
ATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACTGCAGTTATTACTGCA
720
A A
h a
a t
2 2
ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC
721
TACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCACCTCATAAATG
780
B N
g d
l e
1 1
GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTG
781
CCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAAC
840

Fig. 4B

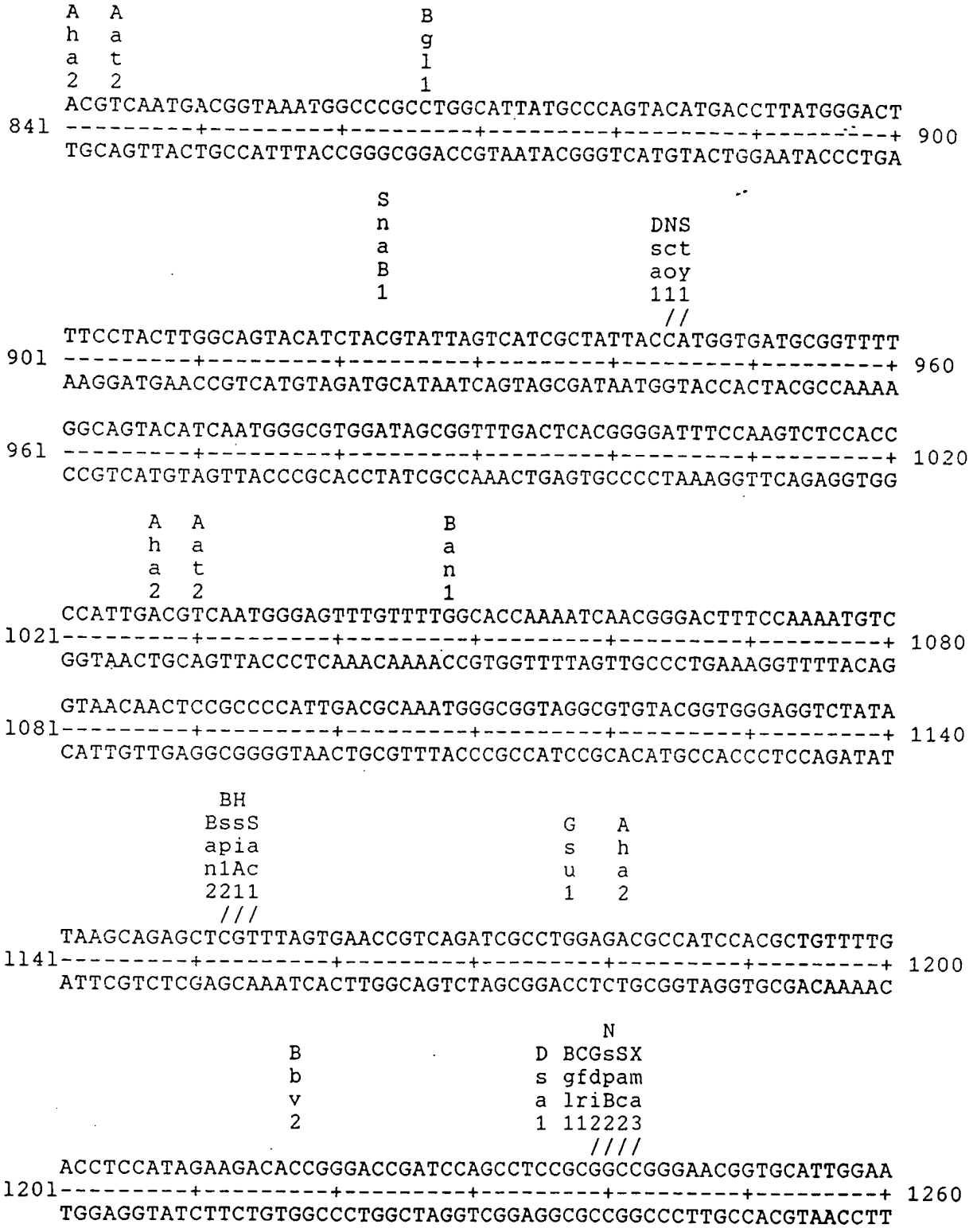

A A
h a
a t
2 2
B
g
l
1
ACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACT
841
TGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATACCCTGA
900
S
n
a
B
1
DNS
sct
aoy
111
//
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT
901
AAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAA
960
GGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC
961
CCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGG
1020
A A
h a
a t
2 2
B
a
n
1
CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC
1021
GGTAACTGCAGTTACCCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAG
1080
GTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA
1081
CATTGTTGAGGCGGGGTAACTGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATAT
1140
BH
BssS
apia
n1Ac
2211
///
G A
s h
u a
1 2
TAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTG
1141
ATTCGTCTCGAGCAAATCACTTGGCAGTCTAGCGGACCTCTGCGGTAGGTGCGACAAAAC
1200
N
B D BCGsSX
b s gfdpam
v a lriBca
2 1 112223
////
ACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGGGAACGGTGCATTGGAA
1201
TGGAGGTATCTTCTGTGGCCCTGGCTAGGTCGGAGGCGCCGGCCCTTGCCACGTAACCTT
1260

Fig. 4C

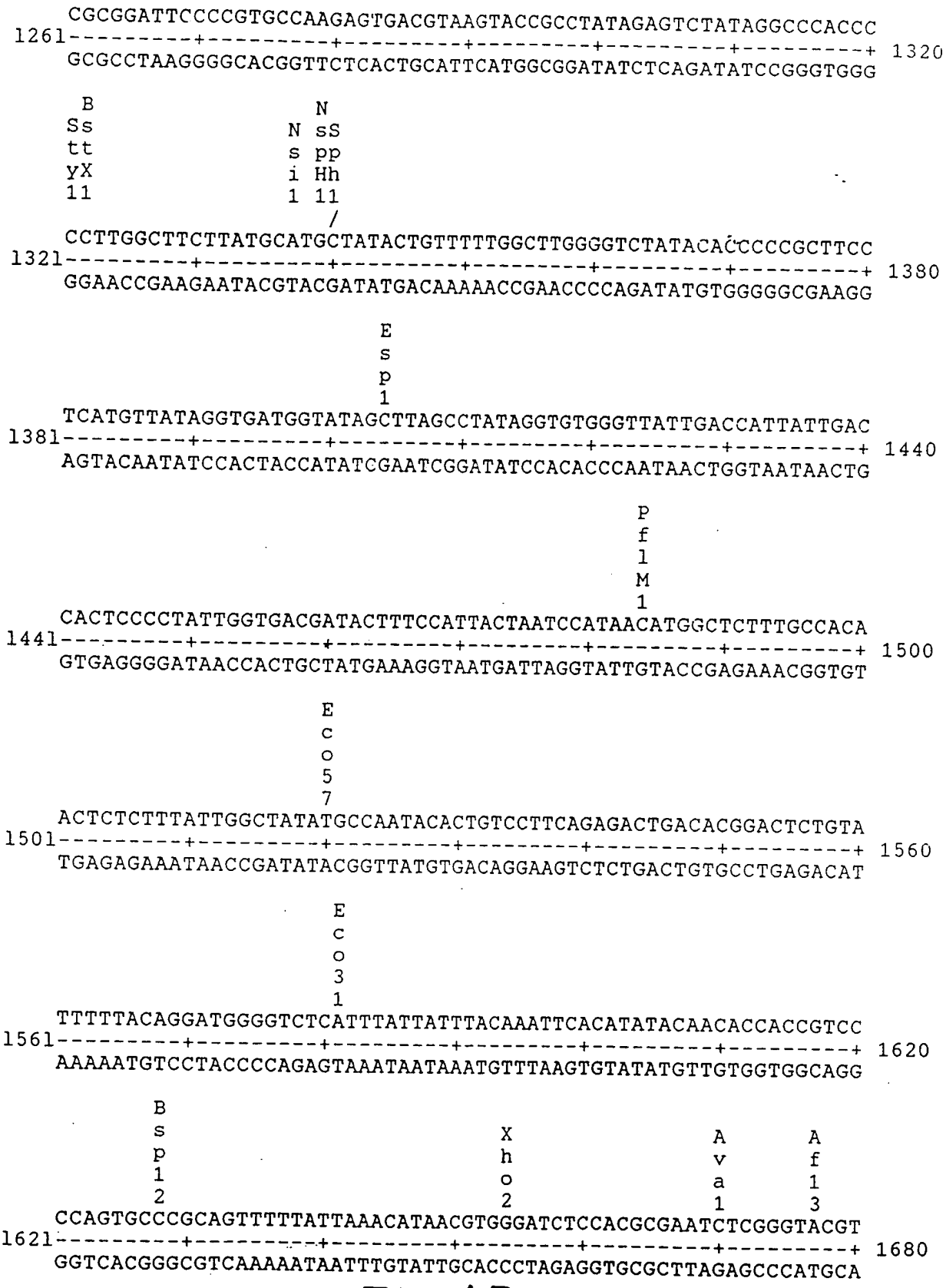

CGCGGATTCCCCGTGCCAAGAGTGACGTAAGTACCGCCTATAGAGTCTATAGGCCCACCC
1261
GCGCCTAAGGGGCACGGTTCTCACTGCATTCATGGCGGATATCTCAGATATCCGGGTGGG
1320
B
Ss
tt
yX
11
N
N sS
s pp
i Hh
1 11
CCTTGGCTTCTTATGCATGCTATACTGTTTTTGGCTTGGGGTCTATACACCCCCGCTTCC
1321
GGAACCGAAGAATACGTACGATATGACAAAAACCGAACCCCAGATATGTGGGGGCGAAGG
1380
E
s
p
1
TCATGTTATAGGTGATGGTATAGCTTAGCCTATAGGTGTGGGTTATTGACCATTATTGAC
1381
AGTACAATATCCACTACCATATCGAATCGGATATCCACACCCAATAACTGGTAATAACTG
1440
P
f
l
M
1
CACTCCCCTATTGGTGACGATACTTTCCATTACTAATCCATAACATGGCTCTTTGCCACA
1441
GTGAGGGGATAACCACTGCTATGAAAGGTAATGATTAGGTATTGTACCGAGAAACGGTGT
1500
E
c
o
5
7
ACTCTCTTTATTGGCTATATGCCAATACACTGTCCTTCAGAGACTGACACGGACTCTGTA
1501
TGAGAGAAATAACCGATATACGGTTATGTGACAGGAAGTCTCTGACTGTGCCTGAGACAT
1560
E
c
o
3
1
TTTTTACAGGATGGGGTCTCATTTATTATTTACAAATTCACATATACAACACCACCGTCC
1561
AAAAATGTCCTACCCCAGAGTAAATAATAAATGTTTAAGTGTATATGTTGTGGTGGCAGG
1620
B
s
p
1
2
X
h
o
2
A
v
a
1
A
f
1
3
CCAGTGCCCGCAGTTTTTATTAAACATAACGTGGGATCTCCACGCGAATCTCGGGTACGT
1621
GGTCACGGGCGTCAAAAATAATTTGTATTGCACCCTAGAGGTGCGCTTAGAGCCCATGCA
1680

Fig. 4D

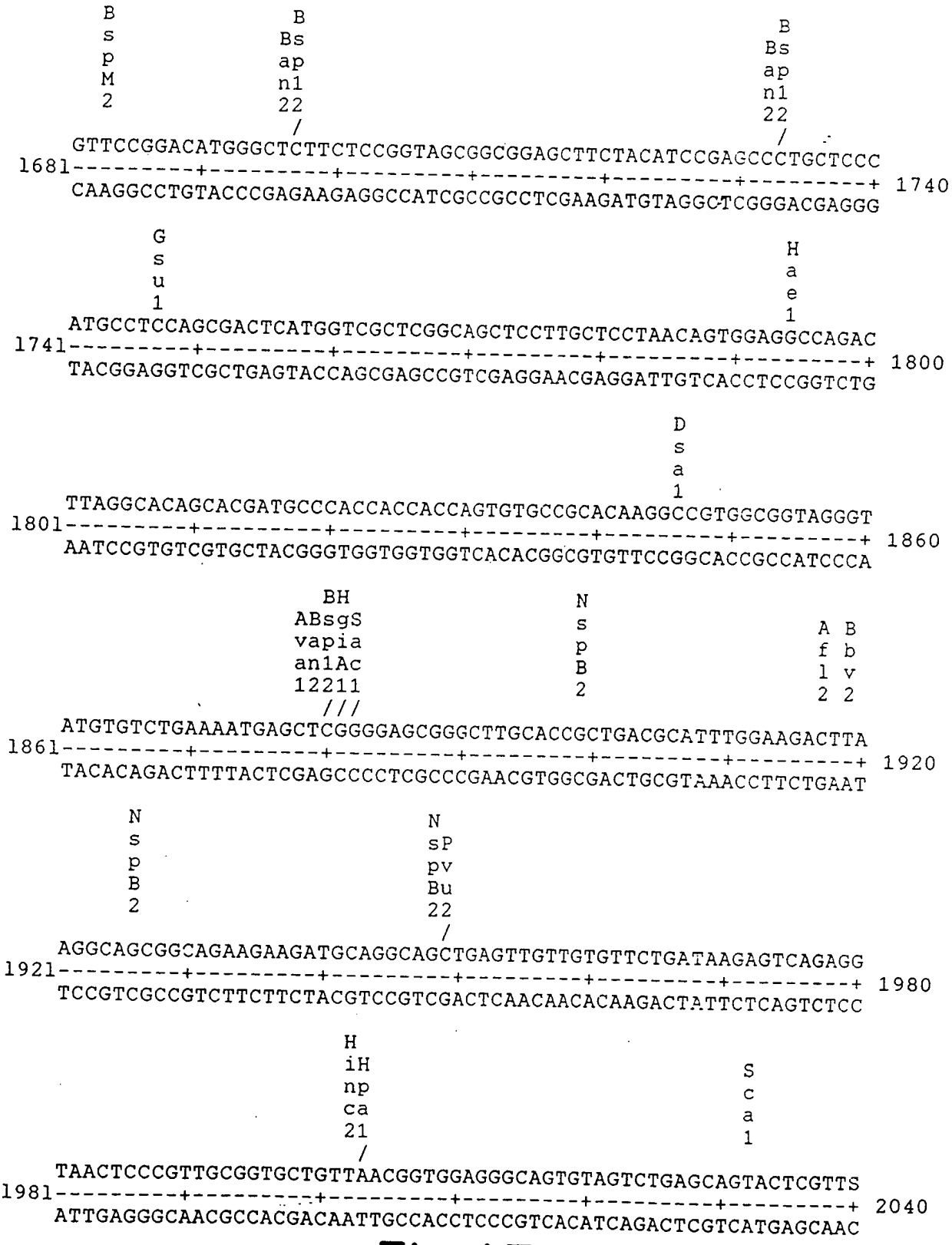

BspM2
BsaI
BsaI
GTTCCGGACATGGGCTCTTCTCCGGTAGCGGCGGAGCTTCTACATCCGAGCCCTGCTCCC
1681
CAAGGCCTGTACCCGAGAAGAGGCCATCGCCGCCTCGAAGATGTAGGCTCGGGACGAGGG
1740
GsuI
HaeI
ATGCCTCCAGCGACTCATGGTCGCTCGGCAGCTCCTTGCTCCTAACAGTGGAGGCCAGAC
1741
TACGGAGGTCGCTGAGTACCAGCGAGCCGTCGAGGAACGAGGATTGTCACCTCCGGTCTG
1800
DsaI
TTAGGCACAGCACGATGCCCACCACCACCAGTGTGCCGCACAAGGCCGTGGCGGTAGGGT
1801
AATCCGTGTCGTGCTACGGGTGGTGGTGGTCACACGGCGTGTTCCGGCACCGCCATCCCA
1860
BHABsgSvapian1Ac12211
NspB2
AflBbvI2 2
ATGTGTCTGAAAATGAGCTCGGGGAGCGGGCTTGCACCGCTGACGCATTTGGAAGACTTA
1861
TACACAGACTTTTACTCGAGCCCCTCGCCCGAACGTGGCGACTGCGTAAACCTTCTGAAT
1920
NspB2
NsPpvBu22
AGGCAGCGGCAGAAGAAGATGCAGGCAGCTGAGTTGTTGTGTTCTGATAAGAGTCAGAGG
1921
TCCGTCGCCGTCTTCTTCTACGTCCGTCGACTCAACAACACAAGACTATTCTCAGTCTCC
1980
HiHnpca21
ScaI
TAACTCCCGTTGCGGTGCTGTTAACGGTGGAGGGCAGTGTAGTCTGAGCAGTACTCGTTS
1981
ATTGAGGGCAACGCCACGACAATTGCCACCTCCCGTCACATCAGACTCGTCATGAGCAAC
2040

Fig. 4E

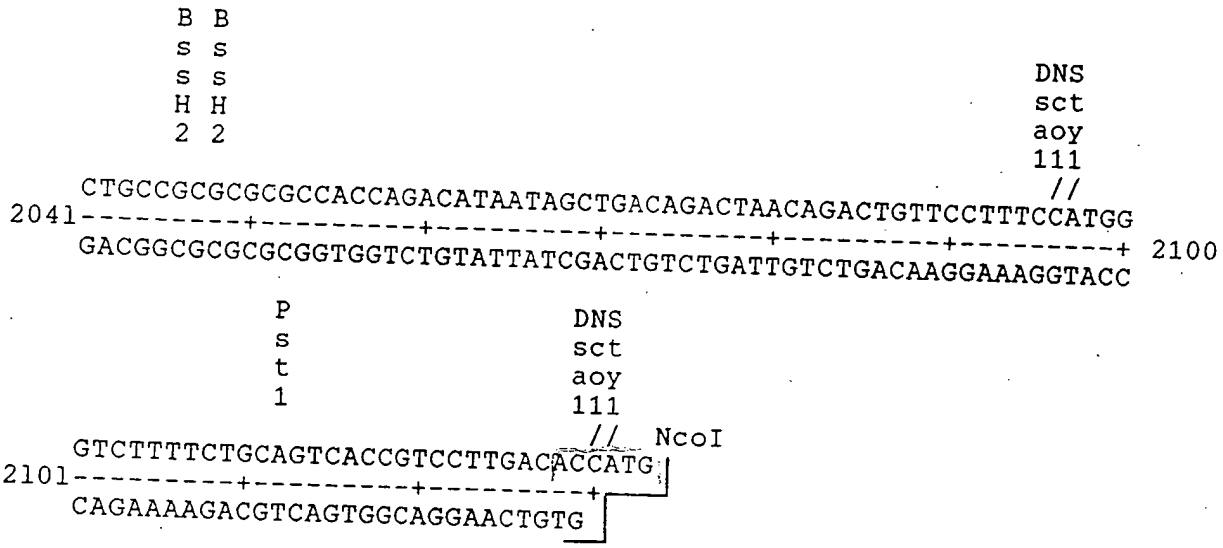

B B
s s
s s
H H
2 2
DNS
sct
aoy
111
//
CTGCCGCGCGCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCCTTTCCATGG
2041
GACGGCGCGCGCGGTGGTCTGTATTATCGACTGTCTGATTGTCTGACAAGGAAAGGTACC
2100
P
s
t
1
DNS
sct
aoy
111
//
NcoI
GTCTTTTCTGCAGTCACCGTCCTTGACACCATG
2101
CAGAAAAGACGTCAGTGGCAGGAACTGTG

Fig. 4F

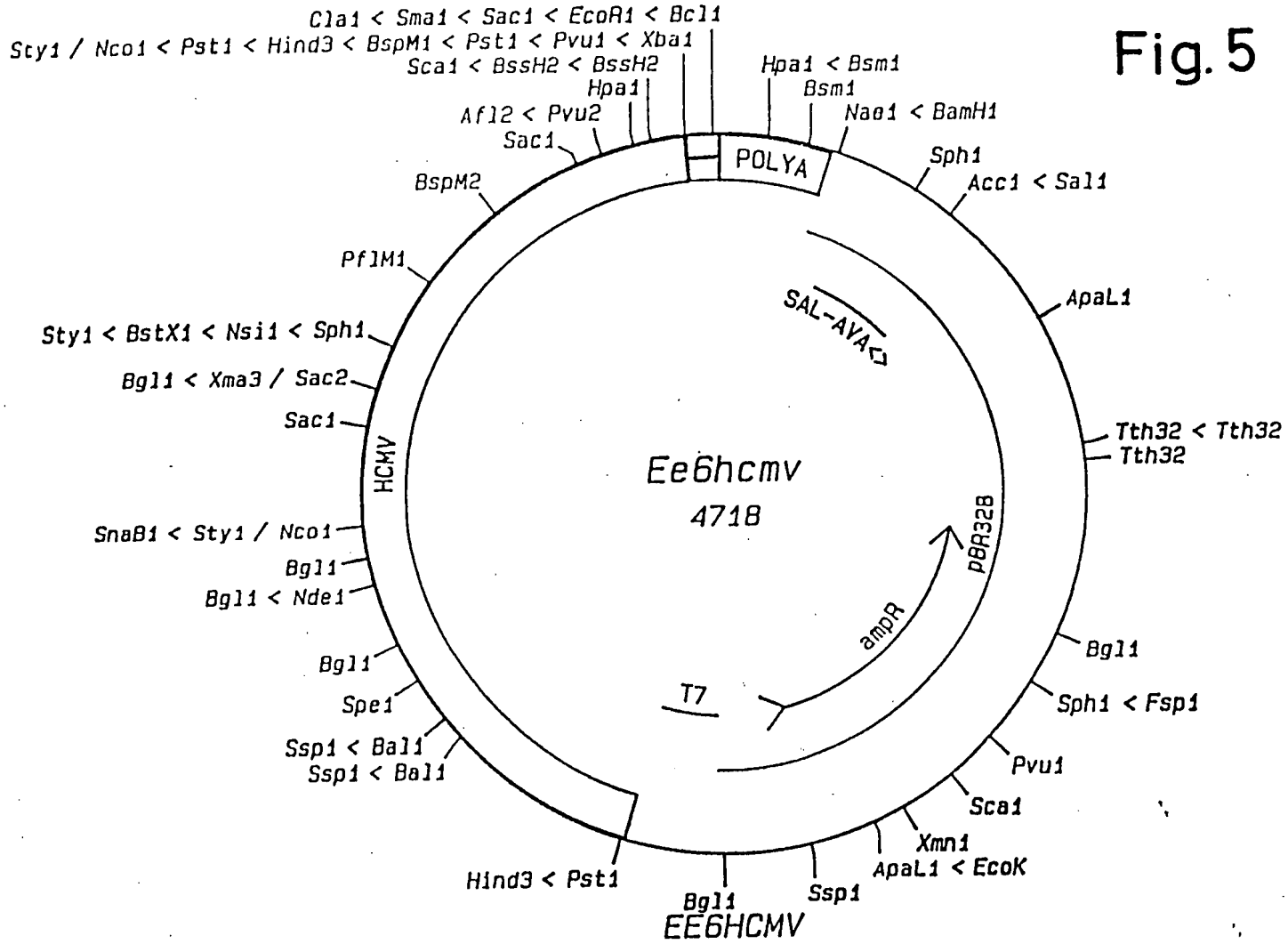


Fig. 5

INSERT HCMV PROMOTER FRAGMENT IN THE HIND3 SITE OF EE6

Mismatch: 0 MinCuts = 1 MaxCuts: 3